

**Nucleotide analogues as reagents for  
site-specific protein-DNA crosslinking**

A thesis submitted for the degree of  
Doctor of Philosophy  
at the  
University of Glasgow

by  
Sharon Shillinglaw Hardie

Division of Molecular Genetics  
Institute of Biomedical and Life Sciences  
Anderson College  
56 Dumbarton Road  
Glasgow

September 2001

© S. S. Hardie, 2001



ProQuest Number: 13818490

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 13818490

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346



12666

COPY 1

The research reported in this thesis is my own and original work except where otherwise stated and has not been submitted for any other degree.



This thesis is dedicated to my mum and dad

Contents	i
Abbreviations	vi
Acknowledgements	viii
Summary	ix

## Chapter 1: Introduction

1.1 The crosslinking of proteins to DNA	1
1.2 Photochemical methods for protein/DNA crosslinking	2
1.3 Phosphorothioates and Azidophenacyl bromide	6
1.4 Cysteines and Azidophenacyl bromide	6
1.5 Sulfur-containing UV crosslinking agents	7
1.6 Chemical methods for protein/DNA crosslinking	7
1.7 Disulfide crosslinking	8
1.8 Crosslinking of Tn3 resolvase	8
1.9 Use of modified oligonucleotides in Tn3 resolvase	10
1.10 Site-specific recombination	11
1.11 Functions of site-specific recombination	12
1.12 The recombinases	13
1.13 The Tn3 transposon	13
1.14 The recombination site, <i>res</i>	14
1.15 Tn3 resolvase	15
1.16 The $\gamma\delta$ -resolvase-site I co-crystal structure	16
1.17 Binding properties of <i>Tn3</i> resolvase	17
1.18 <i>In vitro</i> reaction catalysed by Tn3 resolvase	18
1.19 Formation and functions of the synaptic complex	18
1.20 Models for Strand Exchange	19
1.21 The mechanism/chemistry of strand exchange	21
1.22 Crosslinking methods to study binding/synapsis and strand exchange	23
1.23 Aims of research	24

## **Chapter 2: Synthesis of thiol-containing nucleotide analogues**

2.1 Introduction	26
2.2 Synthesis of the diazirine-containing nucleotide analogue	27
2.3 Thiol-containing pyrimidine compounds	28
2.4 Direct Thioarylation Methods	29
2.5 Indirect Methods of thiol incorporation	31
2.6 Phosphoramidite Synthesis	31
2.7 Synthetic Routes to thiol targets	32
2.8 Experimental	33
2.9 Attempted Synthesis of SH P26	60
2.10 Attempted Synthesis of SH P27	61
2.11 Discussion	61
2.12 Conclusions	67

## **Chapter 3: Materials and Methods**

3.1 Bacterial Strains	69
3.2 Plasmids	69
3.3 Chemical Reagents	69
3.4 Oligonucleotides	70
3.5 Antibiotics	70
3.6 Bacterial growth media	70
3.7 Bacterial Cultures	71
3.8 Preparation of competent <i>E.coli</i> cells	71
3.8.1 Calcium chloride method ( $\text{CaCl}_2$ )	71
3.8.2 Rubidium chloride method ( $\text{RbCl}$ )	72
3.9 Transformation of <i>E.coli</i> with plasmid DNA	72
3.10 Phenol Extraction	72
3.11 Ethanol Precipitation of DNA	73
3.12 Small scale preparation of plasmid DNA	73
3.12.1 Alkaline lysis method	73
3.12.2 'Boiling Prep'	74

3.12.3 Qiagen Miniprep	74
3.13 Large scale preparation of plasmid DNA	74
3.14 UV spectroscopy	76
3.15 Restriction endonuclease digestion of DNA	76
3.16 5' Phosphorylation of DNA	77
3.17 Filling in 3' recessed ends of DNA fragments	77
3.18 3'-end labelling of DNA fragments	77
3.19 Ligation of DNA restriction fragments	78
3.20 Annealing oligonucleotides	78
3.21 Agarose gel electrophoresis	79
3.22 'Single Colony Gel' Analysis	79
3.23 Purification of DNA from low melting point gels	80
3.24 Polyacrylamide Gel Electrophoresis	80
3.24.1 SDS-polyacrylamide gels	80
3.24.2 Non-denaturing polyacrylamide gels	81
3.24.3 Denaturing polyacrylamide gel electrophoresis	82
3.24.4 DNA sequencing gels	83
3.25 DNA synthesis	83
3.26 DNA molecular weight standards	84
3.27 Purification of DNA from polyacrylamide gels	84
3.28 Loading buffers	84
3.29 Gel staining and visualisation	85
3.29.1 Ethidium staining of agarose and polyacrylamide gels	85
3.29.2 Visualisation of radiolabelled DNA	85
3.29.3 Visualisation of proteins	85
3.29.4 Visualisation of oligonucleotides	85
3.30 Tn3 resolvase	86
3.31 <i>In vitro</i> recombination	86
3.32 Band shift analysis	87
3.33 Preparation of competent BL21(DE3)pLysS cells	87
3.34 Expression of the mutants in the expression strain BL21(DE3)pLysS	88

3.35 Purification of cysteine mutants of Tn3 resolvase	88
3.36 Construction of artificial supercoiled substrates	91
3.37 <i>In vivo</i> testing of mutant resolvase plasmids	92
3.38 Synthesis of modified oligonucleotides	93
3.39 Ammonia deprotection of the modified oligonucleotides	93
3.40 $\beta$ -mercaptoethanol deprotection of DNP-containing oligonucleotides	93
3.41 Alkylation of thiol-containing oligonucleotides	94
3.42 Crosslinking reactions	94
3.43 2D gel analysis of crosslinked complexes	95

#### **Chapter 4: Synthesis and characterisation of novel cysteine mutants**

4.1 Introduction	96
4.2 The DNA binding domain	97
4.3 Cysteine Mutants of Tn3 resolvase	99
4.4 Isolation of Novel DNA binding domain mutants	99
4.5 The construction of plasmids encoding cysteine mutants	103
4.6 Cloning of C-terminal sequences into the expression plasmid	104
4.7 <i>In vivo</i> recombination assay	104
4.8 Expression of the mutants in the strain BL21(DE3)pLysS	105
4.9 Purification of cysteine resolvases	105
4.10 Concentration of cysteine mutants	106
4.11 <i>In vitro</i> resolution of pMM5 using novel cysteine mutants	107
4.12 Band Shift analysis of mutants	108
4.13 Binding of resolvase cysteine mutants to <i>res</i> DNA	109
4.14 Binding of resolvase cysteine mutants to site 1 DNA	111
4.15 Discussion	112
4.16 Conclusions	115

## **Chapter 5: Crosslinking using modified oligonucleotides and cysteine mutants**

5.1 Uses of modified oligonucleotides	116
5.2 Oligonucleotides containing the DNP group, T1-T4 and B1-B5	116
5.3 UV Spectrometry of the DNP modified oligonucleotides	116
5.4 Deprotection using $\beta$ -mercaptoethanol and alkylation using IAF	118
5.5 Binding of the modified oligonucleotides to the cysteine mutant proteins	121
5.6 Crosslinking to mutant proteins	122
5.7 Crosslinking of thiol-modified oligonucleotide B1	123
5.8 Crosslinking of thiol-modified oligonucleotide B3	125
5.9 Crosslinking of thiol-modified oligonucleotide B5	125
5.10 Discussion	127

## **Chapter 6: Investigation of the stereochemical properties of Tn3 resolvase**

6.1 Introduction	131
6.2 Construction of modified substrates	132
6.3 Resolution of pMSH2	134
6.4 <i>EcoRV</i> titration of pMSH3	134
6.5 Resolution of pMSH3	135
6.6 Discussion	136
6.7 Future Experiments	137

## **Chapter 7: Further work**

<b>Bibliography</b>	140
---------------------	-----

## Abbreviations

### Units

k =  $10^3$

m =  $10^{-3}$

$\mu$  =  $10^{-6}$

n =  $10^{-9}$

p =  $10^{-12}$

f =  $10^{-15}$

bp = base pairs

kb = kilo base pairs

A = Ampères

V = Volts

Å = Angstroms

Ci = Curies

° C = degrees centigrade

g = grammes

M = molar

m = metres

h = hours

mol = moles

sec = seconds

rpm = revolutions per minute

cps = counts per seconds

cpm = counts per minutes

ppm = parts per million

### Chemicals/Reagents

Ac<sub>2</sub>O = acetic anhydride

AcOH = acetic acid

APS = ammonium persulfate

ATP = adenosine triphosphate

dNTP = deoxynucleoside triphosphate

DMAP = 4-(dimethylamino)pyridine

DCC = *N, N'*-dicyclohexylcarbodiimide

DMF = dimethyl formamide

DNBSCl = 2, 4-dinitrobenzene sulfonyl chloride

DNP = 2, 4-dinitrophenyl

DTT = dithiothreitol

EDTA = ethylenediaminetetraacetic acid (disodium salt)

FDNB - 1-fluoro-2, 4-dinitrobenzene

IAF = 4(5)-(iodoacetamido)-fluorescein  
IPTG = isopropyl- $\beta$ -D-thiogalactopyranoside  
KOAc = potassium acetate  
MeCN = acetonitrile  
NaSH = sodium hydrosulfide  
PMSF = Phenylmethylsulfonyl fluoride  
*sec*-BuLi = secondary butyllithium  
SDS = sodium dodecyl sulfate  
TEMED = *N, N, N', N'*- tetramethylethylenediamine  
THF = tetrahydrofuran  
Tris = *tris*-(hydroxymethyl)aminoethane

#### Other terms

<i>ori</i> = origin of replication	PAGE = polyacrylamide gel electrophoresis
$X^r$ = resistance to X	NMR = nuclear magnetic resonance
IR = infrared	$M^+$ = molecular ion
$R_f$ = relative fraction	mp = melting point
J = coupling constant	$v_{\max}$ = maximum absorption
$m/z$ = mass/charge ratio	tlc = thin layer chromatography
$\epsilon$ = extinction coefficient	$\lambda_{\max}$ = maximum wavelength of absorption
$\delta$ = chemical shift (ppm)	

The abbreviations used in the experimental section (Chapter 2) are routinely used in Organic Chemistry.

Km = resistant to kanamycin	Tc = resistant to tetracycline
Cm = resistant to chloramphenicol	Ap = resistant to ampicillin



## Acknowledgements

Much thanks to my supervisor Marshall Stark for his infinite patience and for the 'speedy' reading of this book, and for forever imprinting "Mary's Prayer" on my memory (Danny Wilson rules o.k). Much appreciation to Martin Boocock for his "pakora-fuelled" chats and endless cups of tea. Thanks to Mary for never losing her cool when all around her were losing theirs, and for making life on the 6th floor great. Thanks to everyone on the 6th floor for their support; Sally, Trish, Alistair (watch out Jaws is about), Elizabeth, 'smiling' Sandra, Kevin, J and Aram (da man) and Amy.

Thanks to the prep. room ladies for making life easier with their sterile media and glassware.

To the chemistry department, thanks to Professor Robins for his supervision. Special thanks to Tony (Mr mass spec.), Kim (Mrs microanalysis), Vikki (Ms IR) for tea and chats and to Jim for the  $^{32}\text{P}$  stuff. Thanks to Douglas Picken for his advice on phosphoramidites. A big thanks to inhabitants past and present; Russell, Graham (ma two front teeth!!), Emma, Vikki, Siobhan, Christine, Stuart, Andy and others.

To my bestest pals in the world Caroline (Curly) and Marnie (Chocolate) for being there in times of need with wine and fun and for generally being fab chicks!! To Drew who has widened my musical taste and shared many nights out in the LB (those were the days!).....

Thanks to my family, Mum, Dad, Trap and Jenny (RIP) and the newest arrival Meg for your love and support throughout the years.

## Summary

A new method for site-specific protein-DNA crosslinking was developed. There is a need for new, improved and more efficient methods because existing strategies generally give very low yield of crosslinked complexes, or are very unselective.

The chosen strategy involved the formation of disulfide bonds between the protein and DNA. This required creation of a thiol at a specific base in the DNA and a cysteine thiol at a specific residue in the protein. Nucleotide analogues with a thiol at the C-5 position of the pyrimidine ring were made as the phosphoramidite derivative. This allowed the thiol-containing analogue to be incorporated into oligonucleotides.

The crosslinking strategy was tested by attempting to crosslink the site-specific recombinase Tn3 resolvase to its DNA binding site. An X-ray structure of Tn3 resolvase bound to DNA has been solved, making it a good model on which to test the method. Cysteine mutants of the Tn3 resolvase DNA binding domain were created, at residues predicted to be close to the modifiable thymidine residues at the binding site. The mutations did not abolish binding and recombination by the protein.

Binding analysis of the cysteine mutants with the oligonucleotides did indicate that the binding affinity of the proteins was reduced. Some crosslinked complexes to the thiol-containing oligonucleotides were observed, albeit in much lower yield than had originally been hoped for.

The stereochemical course of the strand exchange reaction catalysed by Tn3 resolvase was investigated by a strategy involving a phosphorothioate linkage placed at the point of resolvase-catalysed DNA strand cleavage.

## **Chapter 1: Introduction**

## **1.1 The crosslinking of proteins to DNA**

Protein/DNA interactions play a crucial role in many important biological functions. The cellular process of transcription and replication rely on proteins and DNA interacting, forming nucleoprotein complexes. A way of studying protein/DNA interactions is to crosslink two interacting species together. The requirements for the formation of crosslinks between proteins and DNA are that there are chemically reactive sites in the protein and similar sites in the DNA that can come close enough to each other so that a chemical reaction can occur, leading to a covalent bond being formed between the two entities.

The crosslinking of proteins and DNA plays an important role in the analysis of biological processes. There are many different ways of forming covalent bonds between DNA and the proteins that bind to it, and many different ways to analyse the protein-DNA complexes that are formed (reviewed in Welsh and Cantor, 1984; Meisenheimer and Koch, 1997; Williams and Konigsberg, 1991).

Classical protein/protein crosslinking makes use of sites in proteins that are chemically reactive, for example cysteine thiols and lysines which are good nucleophiles. A problem with protein/DNA crosslinking is the lack of functionality within DNA. The amine groups of the bases tend to be chemically unreactive, although most of the bases within DNA have aromatic character and can be activated with UV light at wavelengths of 260 nm.

The first crosslinking between protein and DNA involved irradiation with UV light at 260 nm, which activates the pyrimidine rings, causing a Michael addition reaction between a nucleophilic protein side chain and the pyrimidine ring. It is very difficult to predict the amount and position of crosslinking as many bases in the DNA are activated by this wavelength of light, and protein degradation can occur due to activation of the aromatic amino acids, leading to multiple crosslinks being formed between the protein and DNA.

In some cases the random introduction of a specific type of reactive nucleotide has led to discovery of the sites in DNA that contact the protein (or not). Where there is crystallographic data available for the protein and/or the protein/DNA complex the crosslinking agents can be targeted to the sites which are known to interact, as with the Tn3 resolvase system.

Protein/DNA crosslinking can be analysed using Edman degradation, a technique used to determine the amino acid sequence of proteins. Once the crosslinking has been carried out and the protein/DNA complex is degraded, the point at which the reaction fails is the site where the amino acid is crosslinked to the DNA. Another method used to detect the presence of crosslinking is by using polyacrylamide gel electrophoresis (PAGE). The crosslinking reaction is run out on a gel containing SDS, which will abolish all protein-protein interactions, leaving only crosslinked protein-DNA species visible on the gel. Mass spectrometry can also be used to identify protein-DNA crosslinks. The crosslinked complex is first subjected to proteolysis. The fragment containing the crosslink is identified by PAGE, excised from the gel, and then analysed using electrospray mass spectroscopy (Wong *et al.*, 1998).

Here a new method of inducing covalent, site-specific crosslinking between protein and DNA will be developed. Specific sites in the DNA will be modified with nucleotide analogues containing thiol groups, which will be targeted to engineered cysteine residues in the protein, and disulfides will be formed between the two.

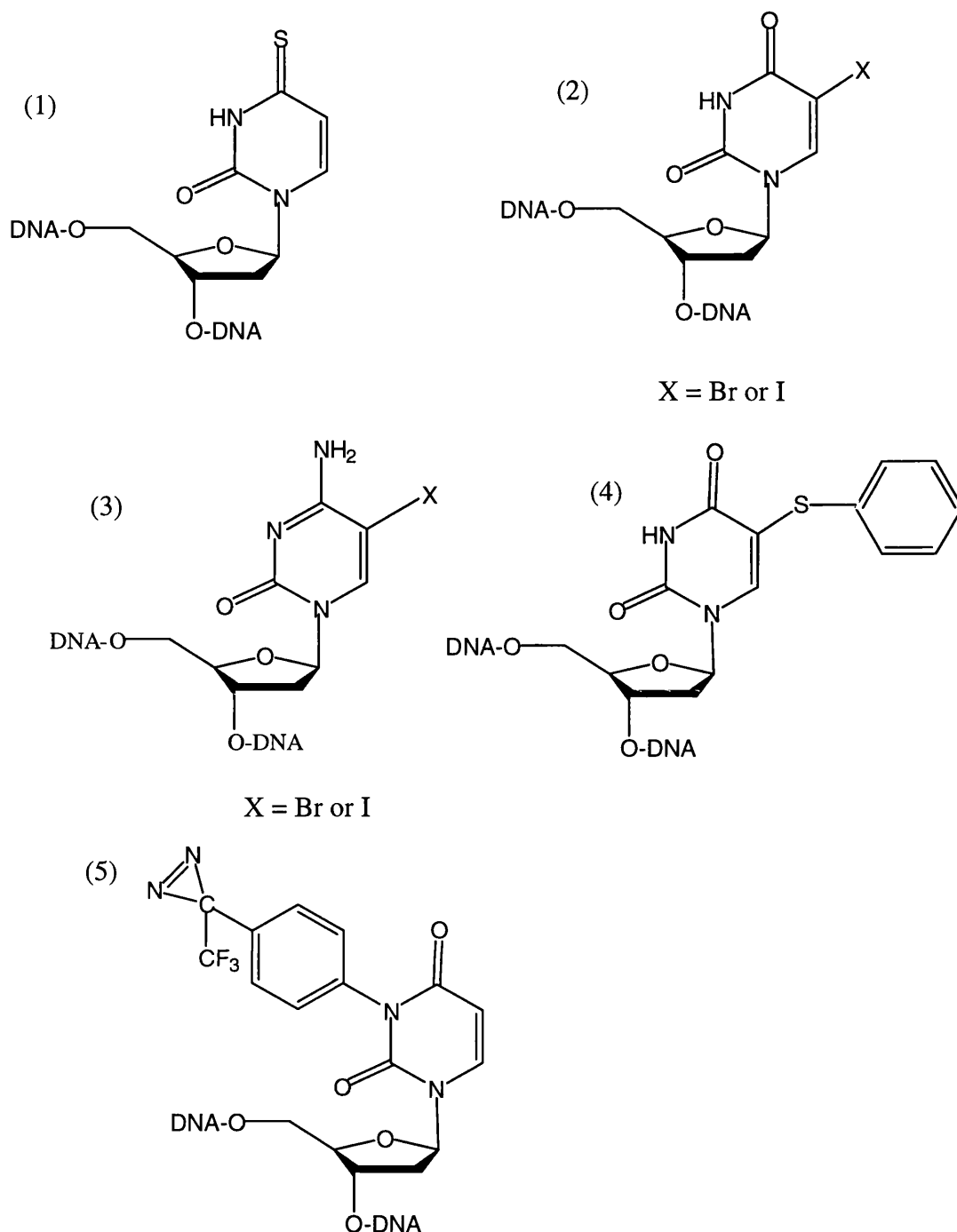
## **1.2 Photochemical methods for protein/DNA crosslinking**

The simplest way to crosslink a protein to DNA is to use UV irradiation of the DNA-protein binding complex. This leads to photochemical activation of the pyrimidine rings such that a nucleophilic amino acid can attack the 5 or 6 position of the ring. Using this method, with enough material and extensive purification of the product, identification of protein-DNA contacts has been made by Edman

degradation. The advantages of using natural DNA for photocrosslinking are that since no modifications need to be made to the DNA, no structural perturbations of the nucleoprotein complex will occur. However the irradiation below 300 nm will lead to absorption by other chromophores in the system, leading to photochemical damage which could lower the yield of crosslinked complex and hinder analysis of the products. Multiple activations will lead to difficulties when trying to identify the site of crosslinking, and both the protein and the nucleic acid need to be sequenced to identify the crosslink. A way to activate the DNA towards crosslinking is to increase the chemical reactivity by introducing functional groups that can be activated easily and with more specificity. Examples of some commonly used photoactivatable nucleotide analogues are shown in Figure 1.1.

Photoactivatable nucleotide analogues such as 4-thiothymidine, azides and halogenated nucleotide analogues have been used, many of which are commercially available. Upon UV irradiation at over 300 nm, these molecules form reactive radical species which will rapidly crosslink to amino acid side chains in the protein that are in close enough proximity to the modified base. Lasers have been used to induce crosslinks and have several advantages over conventional light sources (for review see Pashev *et al.*, 1991) in that the rapid formation of crosslinks will minimise the likelihood that structural rearrangements will occur.

A more selective method is to synthesise an oligonucleotide which contains a functionality at a specific nucleotide that can be activated directly using UV irradiation or can be used to attach a photoreactive agent. This approach reduces purification requirements, since only one crosslinked product is produced. This product is then subject to Edman degradation and the point at which the degradation fails is taken as the point of crosslink formation.



**Figure 1.1 Some photoactivatable nucleotide analogues**

Illustration of some photoactivatable analogues, which have been used for protein-DNA crosslinking. (1) 4-thio-2'-deoxyuridine, (2) 5-bromo/iodo-2'-deoxyuridine, (3) 5-bromo/iodo-2'-deoxycytidine, (4) 5-(phenylthiomethyl)-2'-deoxyuridine and (5) 5-[4-3-(trifluoromethyl)-3H-diazirin-3-yl]phenyl]-2'-deoxyuridine.

5-Bromo-2'-deoxyuridine (5BrdU; Figure 1.1, compound 2) is an analogue of thymidine that is activated at 308 nm, thus photochemical damage is minimised. In addition 5BrdU is very similar in size to the thymidine C-5 methyl group which has a Van der Waals radius of 2.0 Å; the 5BrdU radius is 1.95 Å, making it a good isosteric analogue. Therefore there are minimal structural perturbations to protein-DNA complexes and the binding affinities of DNA containing 5-BrdU are similar to the natural DNA. Incorporation of 5BrdU *in vivo* into the DNA has also been used. The analogue is a substrate for the polymerase enzymes and can be incorporated at multiple sites in the DNA, leading to higher yields of crosslinking upon UV irradiation. For example, the archaeobacterial chromosomal protein MC1 was crosslinked using 5BrdU-enriched DNA (Katouzian-Safadi *et al.*, 1991), as was the *lac* repressor (Barbier *et al.*, 1984.) Protein-DNA contacts have been identified using crosslinking with 5BrdU. The oncogene product Myc is a sequence-specific DNA-binding protein. DNA fragments were systematically modified with 5BrdU and crosslinking identified regions of the protein that came into close contact with the DNA by tryptic digest followed by peptide sequencing (Dong *et al.*, 1994). Points of contact between the Rel transcription factors and its target DNA were identified using 5BrdU crosslinking (Liu *et al.*, 1994). The bacteriophage R17 coat protein was crosslinked using 5-bromouracil, and the products were analysed by SDS PAGE (Willis *et al.*, 1994).

5-Iodo-2'-deoxyuridine (5IdU; Figure 1.1, compound 2) is another analogue of thymidine that can be activated by UV irradiation at 308 nm. In addition, 5IdU is very similar in size to natural thymidine, with a Van der Waals radius for the iodine atom of 2.15 Å (see above). 5-Iodouracil has been used to crosslink both RNA and DNA to proteins (Willis *et al.*, 1993) generally with greater efficiency than the equivalent brominated analogue. A yield of 40% crosslinking of bacteriophage R17 coat protein to its substrate DNA was reported. 5IdU was used to crosslink DNA-*EcoRI* DNA methyltransferase complexes (Wong *et al.*, 1998). High yields of 60% crosslinked complex were obtained, and the products were further analysed using electrospray mass spectrometry. The MutS DNA mismatch



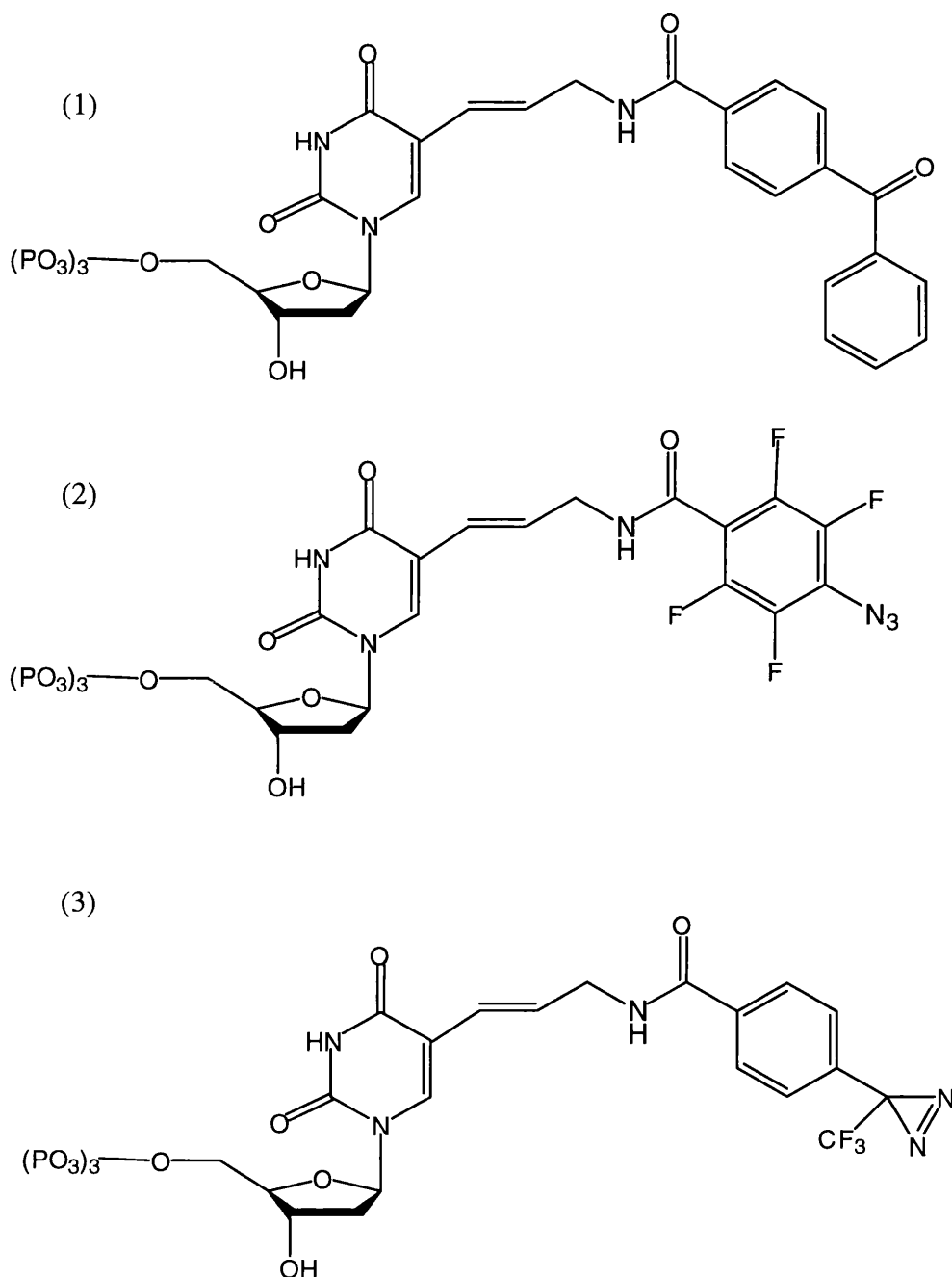
repair protein was crosslinked to heteroduplex DNA containing 5-iodo-2'-deoxyuridine (Malkov *et al.*, 1997). A technique was developed that used 5IdU in RNA ligands to identify high affinity proteins that bind to these ligands (Jensen *et al.*, 1995).

Vaccinia topoisomerase I enzyme binds and cleaves duplex DNA at sites containing a specific recognition sequence. The enzyme was crosslinked to its target DNA, using 5-bromo-2'-deoxycytidine (5BrdC; Figure 1.1, compound 3) placed at the specific recognition sequence of the enzyme. The crosslinking efficiency was 15%. Following isolation and sequencing of peptide-DNA photoadducts it was possible to propose a model for the protein-DNA interface, and in conjunction with the crystal structure of the N-terminal fragment to speculate on the possible roles of important residues at the interface (Sekiguchi *et al.*, 1996).

A photoreactive diazirine analogue 5-[4-(3-(trifluoromethyl)-3H-diazirin-3-yl)phenyl]-2'-deoxyuridine (Figure 1.1, compound 5) was crosslinked to *EcoRII* and *MvaI* restriction/modification enzymes (Topin *et al.*, 1998).

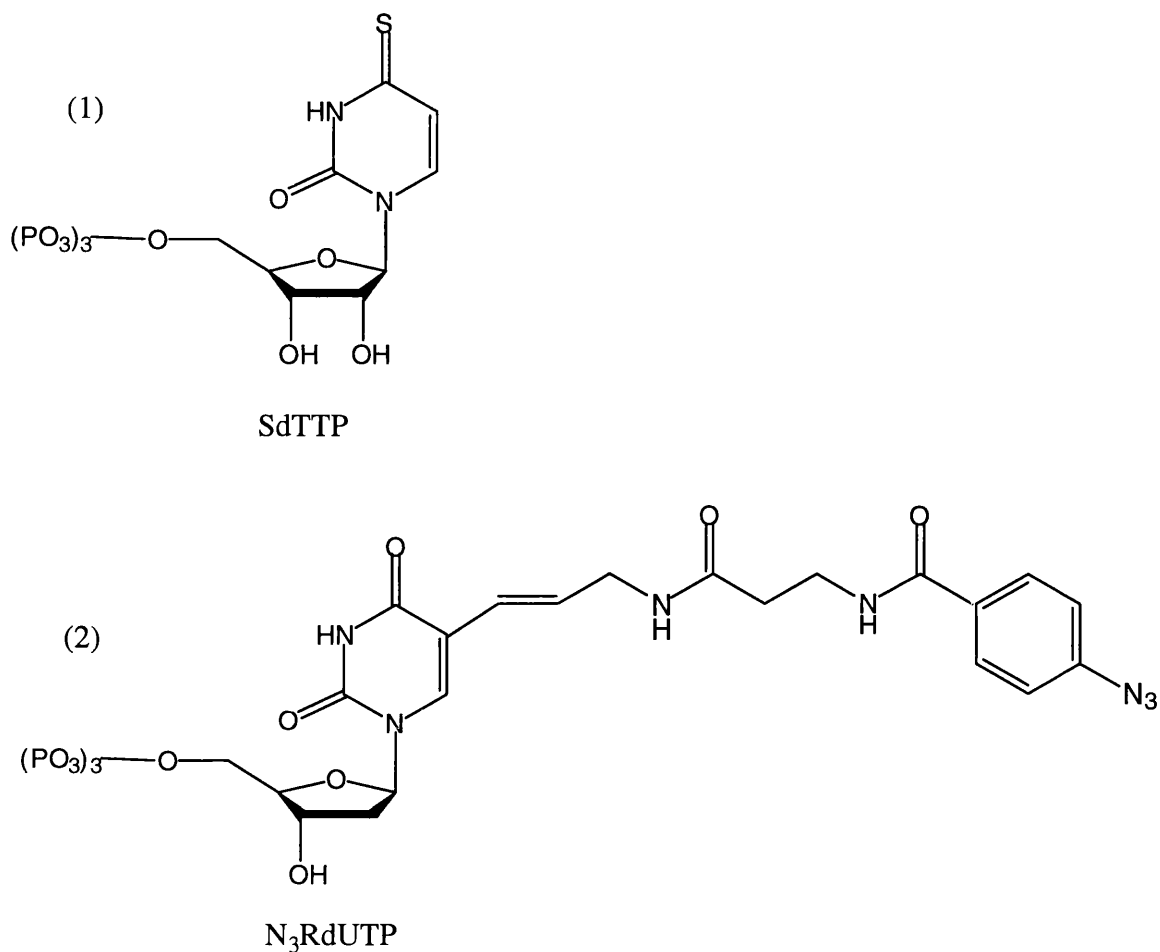
Transcription complexes of yeast RNA polymerase III were studied using four different photoreactive reagents (Tate *et al.*, 1998). These included aryl azide, benzophenone (Figure 1.2, compound 1), perfluorinated aryl azide (Figure 1.2, compound 2) and diazirine (Figure 1.1, compound 3). The last two reagents giving the highest levels of crosslinking.

Yeast transcription complexes were assembled and crosslinked using the photoactive 4-thiodeoxythymidine 5'-triphosphate (S-dTTP; Figure 1.3, compound 1) (Bartholomew *et al.*, 1994) and 5-[*N*-(*p*-azidobenzoyl)-3-aminoallyl]-deoxyuridine triphosphate (N<sub>3</sub>RdUTP; Figure 1.3, compound 2). This reagent resulted in higher yields of crosslinked complexes, 40-45%, compared to 5IdU and 5BrdU (Bartholomew *et al.*, 1990). This reagent was also used to study



**Figure 1.2 Longer distance photoreactive analogues**

These longer length crosslinkers were used to crosslink transcription complexes. A comparative study using the above modified nucleotides showed that the choice of photoreactive group used in photoaffinity labelling is crucial for identifying protein-DNA contacts (Tate *et al.*, 1998). (1) 5-[N-(4-benzoylphenyl)-3-aminoallyl]-dUTP, (2) 5-[N-(4-azido-2,3,5,6-tetrafluorobenzoyl)-3-aminoallyl]-dUTP and (3) 5-[N-(4-(3-(trifluoromethyl)diazirin-3-yl)benzoyl)-3-aminoallyl]-dUTP.



**Figure 1.3 Photoreactive analogues that were used to study yeast transcription complexes**

(1) is 4-thiodeoxythymidine 5'-triphosphate (S-dTTP) and (2) is 5-[N-(p-azidobenzoyl)-3-aminoallyl]-deoxyuridine triphosphate (N<sub>3</sub>RdUTP). Within the transcription complexes more crosslinking was obtained using the 'zero-distance' SdTTP than with the longer reaching N<sub>3</sub>RdUMP (Batholomew *et al.*, 1990 and 1994).

the tracking of viral, bacterial and eukaryotic proteins along DNA (Tinker *et al.*, 1994).

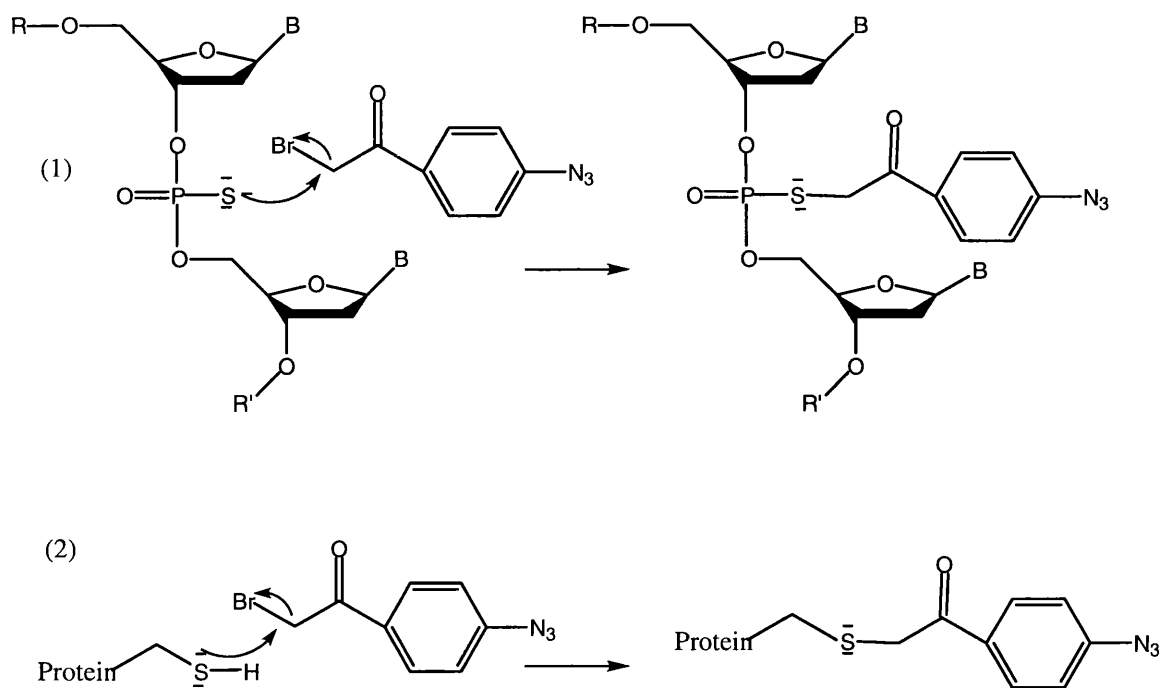
### **1.3 Phosphorothioates and Azidophenacyl bromide**

This methodology involves the synthesis of DNA substrates containing phosphorothioates at specific positions in the backbone. Phosphorothioates have a sulfur atom replacing the non-bridging oxygen in the DNA backbone (see Chapter 6). The phosphorothioate was then modified with azidophenacyl bromide, that places a photoactivatable azide group in the DNA (Figure 1.4.1). This method was used to crosslink Integration Host Factor (IHF) to its target DNA (Yang and Nash, 1994), TATA nucleoprotein complex to DNA (Lagrange *et al.*, 1996), Human T-cell Leukemia Virus Type I Tax protein (Kimzey and Dynan, 1998), bacterial ribonuclease P (Harris *et al.*, 1994) and bacterial RNA polymerase and its promoter (Naryshkin *et al.*, 2000).

Phosphorothioates within oligonucleotides have been crosslinked using platinum complexes (Figure 1.5). These reagents have been used to crosslink complementary oligonucleotides together (Gruff and Orgel, 1991) and transcription factors to their target DNA (Chu and Orgel, 1992).

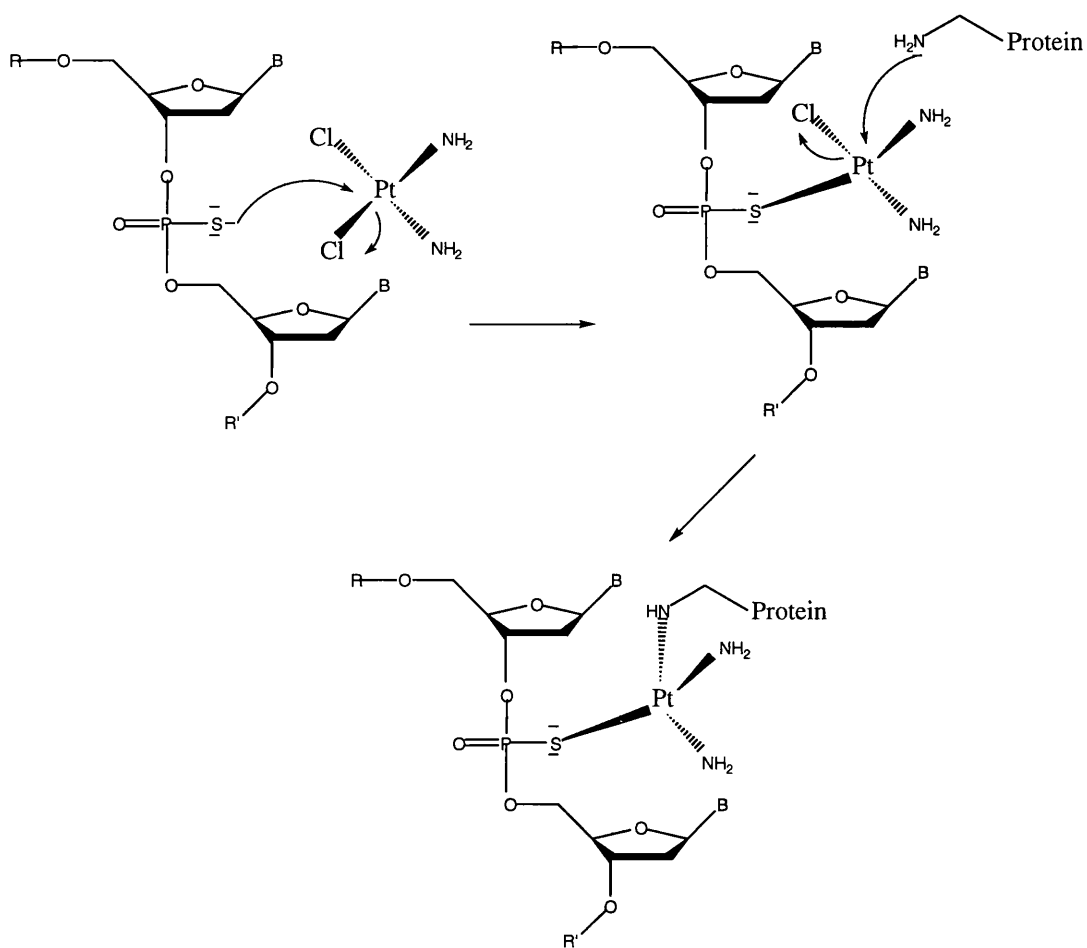
### **1.4 Cysteines and Azidophenacyl bromide**

A way to modify proteins with crosslinking agents is to introduce cysteine residues at specific regions of the protein, modify these cysteines with azidophenacyl bromide, then irradiate this with UV light to induce crosslinks to the target DNA (Figure 1.4.2). This technique has been used to study the LexA DNA-binding motif on operator DNA (Dumoulin *et al.*, 1993), bacteriophage  $\lambda$  Cro (Chen and Ebright, 1993), histone H2A (Lee and Hayes, 1997), LexA Repressor (Dumoulin *et al.*, 1996), transcription activator protein Oct-1 (Cleary *et al.*, 1997) and CAP (Pendergrast *et al.*, 1992).



**Figure 1.4 Azidophenacyl bromide modifications**

In (1) a phosphorothioate in the DNA is modified with azidophenacyl bromide and (2) a protein cysteine thiol is alkylated.



**Figure 1.5 Cisplatin crosslinking**

Illustration of cisplatin crosslinking of a DNA phosphorothioate and a protein lysine residue (cysteine thiols also react).

## 1.5 Sulfur-containing UV crosslinking agents

4-thio-2'-deoxyuridine (4SdU; Figure 1.1, compound 1) has the oxygen atom at position 4 of the pyrimidine ring replaced with sulfur, rendering it photoactivatable with UV light above 300 nm. 4SdU has been incorporated into DNA substrates for *EcoRV* endonuclease and *EcoRV* methyltransferase enzymes (Nikiforov and Connolly, 1992). Combinations of different thymidines in the GATATC target sequence for *EcoRV* were replaced with 4SdU and crosslinking was induced using UV irradiation at 310 nm. Yields of only 6% were accomplished. An RNA phosphoramidite of 4SU that has the 2'-OH protected with a 1-(2-fluorophenyl)-4-methoxypiperidin-4-yl function was used to crosslink the Rev protein of HIV-1 (McGregor *et al.*, 1996).

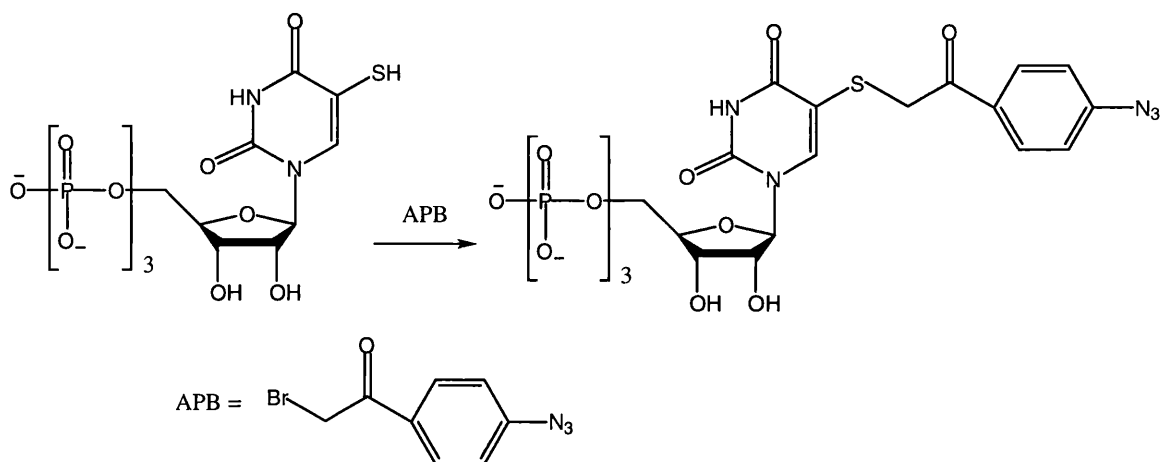
The analogue 5-(phenylthiomethyl)-2'-deoxyuridine (Figure 1.1.4) is a photoreactive nucleobase that is activated at 260 nm (Romieu *et al.*, 2000). It cannot be used for the specific crosslinking because it is activated at 260 nm.

5-Mercaptouridine triphosphate/azidophenacyl (5-SH UTP/APAS; Figure 1.6) was incorporated into RNA utilising *E.coli* and T7 RNA polymerases and yeast RNA polymerases I and III. These analogues can be transcribed into segments of DNA in a site-specific manner. The nucleotide analogue 5-[(4-azidophenacyl)thio]uridine 5'-triphosphate was synthesised (Hanna *et al.*, 1989) and used to study RNA polymerases. 5-mercapto-UTP (5-SH-UTP) has been used to study protein-RNA interactions (He *et al.*, 1995). The same methodology was used to synthesise 5-[(4-azidophenacyl)thio] cytidine 5'-triphosphate (Hanna *et al.*, 1993), also shown to be useful for crosslinking to *E.coli* and T7 RNA polymerases.

## 1.6 Chemical methods for protein/DNA crosslinking

### Non-specific chemical cross-linking

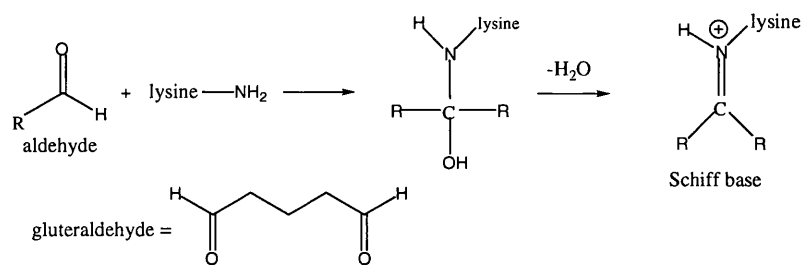
Formaldehyde and glutaraldehyde are examples of "non-specific" chemical crosslinkers. Formaldehyde reacts with several amino acid side chains producing



**Figure 1.6 Nucleotide analogues that are substrates for polymerase enzymes**

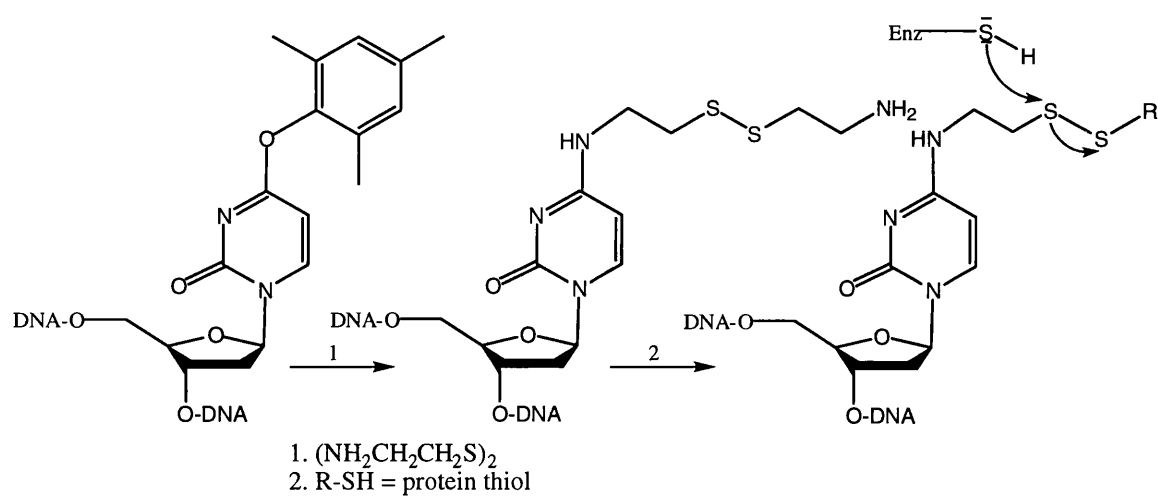
Illustration of 5-mercapto-UTP (5-SH-UTP) and 5-[(4azidophenacyl)thio]uridine 5'-triphosphate 5 [4-APAS]UTP. Used to crosslink *E. coli* and T7 polymerases (Hanna *et al.*, 1989).





**Figure 1.7 Crosslinking with dialdehydes**

Illustration of crosslinking between an aldehyde (e.g. glutaraldehyde) and the primary amine group of lysine, forming a Schiff base crosslink. This method has been used to crosslink the RNA polymerase-*lacUV5* promoter complex (McGhee and von Hippel, 1975).



**Figure 1.8 Illustration of disulfide bond formation in nucleotide analogues**

The synthesis of the thiol tether (step 1), and the reaction that forms the disulfide between a cysteine in HIV-1 reverse transcriptase and the thiol tether in the template DNA (step 2).

a methylol (hydroxymethyl) derivative which can then react further with DNA at amino and imino groups. Glutaraldehyde is a bifunctional aldehyde that forms a Schiff base with primary amines (Figure 1.7).

### **1.7 Disulfide crosslinking**

The convertible nucleotide analogue approach has been used to successfully incorporate thiol-containing bases at specific positions within oligonucleotides (MacMillan and Verdine, 1991). The oligonucleotide is site-specifically modified to carry a tethered nucleophile. The nucleophile is then reacted with a ligand electrophile, usually an amine or a thiol disulfide (see Figure 1.8). The above approach was used to crosslink HIV-1 reverse transcriptase to its template DNA *via* a disulfide (Huang *et al.*, 1998). The cysteine thiol formed a disulfide with a thiol group in the minor groove of the DNA (activated as a mixed disulfide), which was tethered to the N-2 of a dG in the template:primer. The positions of the thiols were systematically varied in both protein and template. Defined cycles of polymerisation were carried out with the reacting thiol partners such that different combinations of dNTPs and ddNTPs moved the enzyme a fixed distance along its substrate. When the DNA template was positioned so that the thiol tether and the engineered cysteine were close enough, a disulfide crosslink was formed, that prevents dissociation of the product, yielding a stalled complex. Disulfide crosslinking has been used to study DNA hairpins (Glick, 1991) that were stabilised by introducing disulfide crosslinks at the terminal bases.

Other reagents that have created a disulfide bond between an oligonucleotide and an enzyme (Chu and Orgel, 1988) have involved modification of the oligonucleotide at the 5' end of the oligonucleotide, which is not suitable for the study of Tn3 resolvase.

### **1.8 Crosslinking of Tn3 resolvase**

To date the crosslinking agents that have been used to study protein-DNA interactions have had limited efficiency and thus usefulness. With most

photoactivatable analogues low yields of crosslinked complex are usually obtained, thus lots of material is required in the reactions. For the study of Tn3 resolvase and other proteins that interact with segments of DNA, e.g. transcription factors or replication machinery, it is important that the modifications do not cause any structural perturbations in the system. One of the most popular crosslinking agents, azidophenacyl bromide, introduces a bulky group at the site of crosslinking, so efficiency of binding of proteins would be effected. Thiols placed at the C-5 position of a pyrimidine (as a thymidine analogue) cause minimal structural perturbations, and so should not affect protein binding. 4SdU and 5IdU are commercially available nucleotide analogues that have been used previously to crosslink Tn3 resolvase to DNA (McIlwraith, 1995).

4-Thiothymidine (Section 1.5; Figure 1.1, compound 1) has also been incorporated into DNA substrates for Tn3 resolvase (McIlwraith, 1995) with high levels of crosslinking to linear DNA of 13%. It was however found to be unsuitable for the intended experiments, as the structure of the DNA duplex was altered in the supercoiled substrate. The O-4 is required for the formation of the Watson and Crick base pairings that are critical for maintaining the helix in the correct conformation. The replacement of this oxygen with sulfur caused changes in the helix structure that affected the binding of the resolvase. Other photoreactive bases, including azidophenacyl bromide could also not be used to study Tn3 resolvase due to the introduction of a sterically bulky group which would cause unnatural conformations of the DNA duplex and so disrupt resolvase binding and synaptic complex formation.

Replacement of thymine with 5-iodouracil (Section 1.2; Figure 1.1, compound 2) has previously been used to study the interactions between Tn3 resolvase and its substrate DNA. Photocrosslinking experiments have showed that at least one of the four resolvase monomers bound at the crossover site must remain physically attached to the same DNA half site before and after recombination (McIlwraith *et*

*al.*, 1996). The crosslinking utilised 5IdU, substituted for thymidines in the TGT binding motif in a synthetic *res* subsite I (Figure 1.9.2), incorporated into a supercoiled substrate, and targeted towards Tyr176 in the resolvase recognition helix. The crosslinking was between the phenolic ring of tyrosine and the uridyl radical. The best results obtained using this reagent yielded crosslinked product between 7 and 15% of the total substrate. However, for some useful mechanistic studies to be done would require more than one modification, so the initial yield would have to be as high as possible.

Convertible nucleotide analogue strategy for disulfide crosslinking (Section 1.7) is not suitable for crosslinking Tn3 resolvase to its target DNA. Although these analogues can be positioned in a site-specific manner, the thiol group is at the end of a two carbon tether. The strategy used here prefers the thiol to be similar to the 5BrU analogues, that is, not too structurally different from the natural nucleotides. The longer tether would place give the thiols a greater degree of flexibility but may cause structural perturbations in the resolvase/DNA complex. The crosslinking agents developed in this study will be tested in the Tn3 resolvase/DNA system. A co-crystal structure of the closely related  $\gamma\delta$  resolvase (Yang and Steitz, 1995) enables us to look closely at the points of interaction between the protein and DNA and so predict possible points of contact that could be crosslinked together. Tn3 resolvase is coded by the transposon Tn3 and carries out a process called site-specific recombination (Section 1.10).

### **1.9 Use of modified oligonucleotides in Tn3 resolvase**

Modified oligonucleotides for all of the *res* subsites can be made and incorporated into supercoiled plasmids (Section 3.36), the natural substrate for Tn3 resolvase. This enables site-specific replacement of bases in the resolvase binding sites with photoactivatable or thiol-containing nucleotide analogues. Modified bases can be incorporated into oligonucleotides, that can be synthesised by the phosphoramidite method (Section 3.25). In order to determine the functions of the individual resolvase subunits, a new method of making covalent

bonds (crosslinks) between one or more subunits and the DNA at specific sequence positions was developed (Chapter 5). Unlike generic 'crosslinking agents' the crosslinking approach used here targets sites in the DNA that are known to interact with the protein, rendering them 'affinity reagents'.

To perturb the catalytic steps involved in strand exchange, a phosphate in the DNA backbone will be made chiral, at the site where the strand is broken by Tn3 resolvase (Figure 1.9.2; Chapter 6).

### **1.10 Site-specific recombination**

Site-specific recombination is a process where DNA strands are cut enzymatically, and rejoined in a different configuration. Tn3 is a bacterial transposon that uses site-specific recombination. Transposons are DNA sequences that can jump from their original site on the genome to other sites within it. These 'jumping' genes allow for the horizontal transfer of genetic information.

Site-specific recombination of DNA plays a major role in many biological processes in prokaryotes and eukaryotes. During conservative site-specific recombination a DNA molecule is cut in both strands at two specific positions and the ends are rejoined to new partners in a series of precise catalytic steps. During the process there is no DNA synthesis or degradation involved, or hydrolysis of the phosphodiester bonds. Site-specific recombination differs from homologous recombination in that extensive homology within the region of strand exchange is not required. The only sequence homology required is at the site where the DNA is broken, i.e. the "overlap" region.

All site-specific recombination systems require a recombinase enzyme that binds to a specific site in the DNA, cleaves it and exchanges both strands in a controlled way to give a recombinant configuration. There are many different types of

recombination systems that require one (or more) recombinase enzyme(s) and specific binding sites for these enzymes on one (or more) pieces of DNA.

### **1.11 Functions of site-specific recombination**

The functions of site-specific recombination vary in nature. Many organisms have one or more recombination systems. These systems are also associated with phages and plasmids. The recombinational apparatus of phage  $\lambda$  controls the integration and excision of the phage to and from the host chromosome. Thus recombination plays an essential role in the life cycle of the phage. In the transposon Tn3, the process of recombination resolves the co-integrate DNA intermediate structures that arise as a result of the transposition of the transposon to another DNA site (Section 1.13).

In yeast, the copy number of the 2 $\mu$  plasmid is regulated by the recombination reactions of the Flp recombinase which inverts a segment of DNA resulting in switching of the direction of replication fork in that region. The invertases Cre of the phage P1 and Hin from phage Mu both use recombination to alter the expression of genes. The Hin invertase of phage Mu has almost 30% homology with Tn3 resolvase. In response to the host environment, recombination occurs inverting the segment of DNA that carries promoters for expressing surface antigens, which will help the bacteria to evade host factors. XerC/D recombination systems in *E.coli* aid DNA stability by maintaining circular DNA replicons in a monomeric state. In humans, V(D)J recombination creates diversity of the immunoglobulin genes.

In *S. cerevisiae* and *D. Melanogaster* sequences corresponding to integrated proviruses have been found, the *Ty* elements of yeast and *copia* elements from flies have coding sequences with homology to reverse transcriptase which mobilise *via* an RNA form but have no infectious properties. Another important transposable element in *D. Melanogaster* is the P element that determines the fertility in progeny. Males carrying P elements crossed with females lacking the

elements have progeny which are infertile (hybrid dysgenesis). The LINES (long interspersed sequences) are derived from transcripts of RNA polymerase II and are found on multiple sites throughout mammalian genomes. LINES L1 belongs to the viral superfamily of retrotransposons like *Ty* and *copia*. Mammalian genomes also contain SINES (short interspersed sequences) that are from the non-viral class of retrotransposons and are derived from transcripts of RNA polymerase III. Site-specific recombination shares many common features with other processes where DNA strands are cut and rejoined at specific sites, especially transposition.

The work in this thesis focuses on components of the bacterial transposon Tn3 and the mechanism of the proteins that are coded by it.

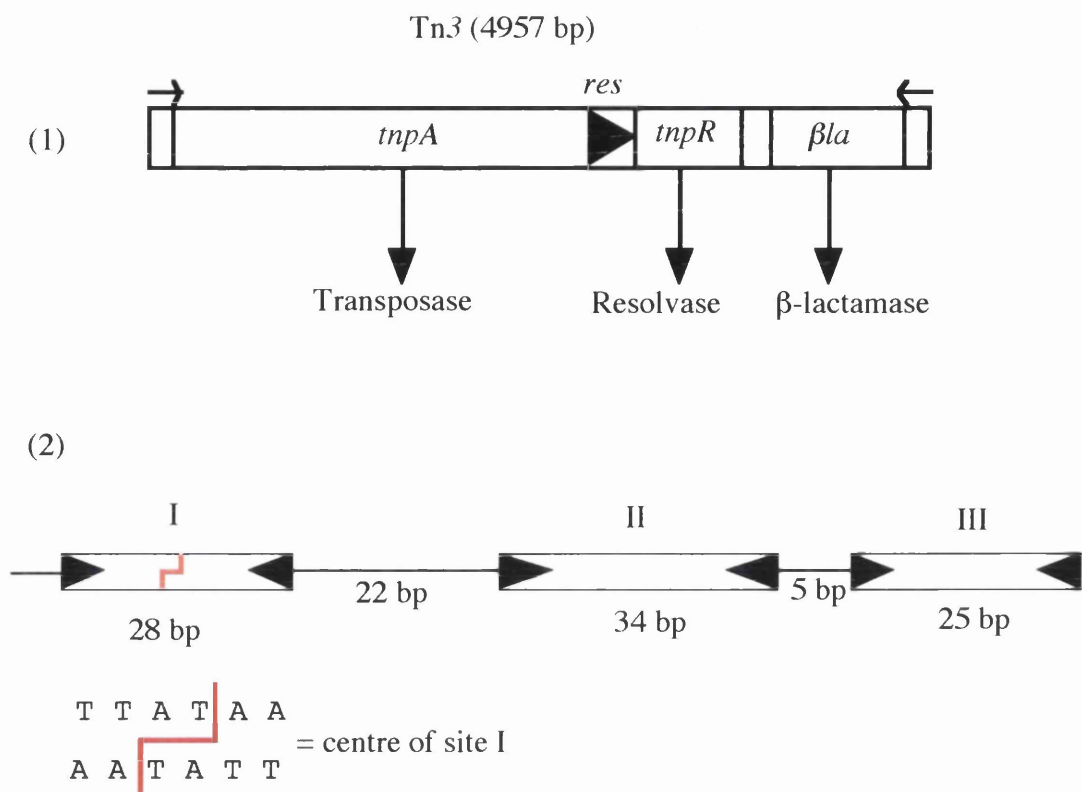
### **1.12 The recombinases**

Two distinct families classified by amino acid sequence and by similarities in their recombination enzymes and recombination sites are the integrase and the resolvase/invertase family. The integrases family known as the 'tyrosine recombinases' have a catalytic tyrosine residue which attacks the DNA phosphodiester backbone, whereas the resolvase family, the 'serine recombinases' have a catalytic serine. In the resolvase family all four strands are broken before the strand exchange takes place, whereas in the integrase family sequential single strand cleavage occurs, the first two cleavages giving rise to formation of a Holliday junction intermediate, which is then resolved into the recombinant configuration after the second pair of strand exchanges (Reviewed in Stark *et al.*, 1992). Sometimes proteins and/or binding sites within the same family are interchangeable. The  $\gamma\delta$  resolvase can act at Tn3 binding sites and *vice versa*. Tn1000 ( $\gamma\delta$ ) shares 90% sequence homology with Tn3.

### **1.13 The Tn3 transposon**

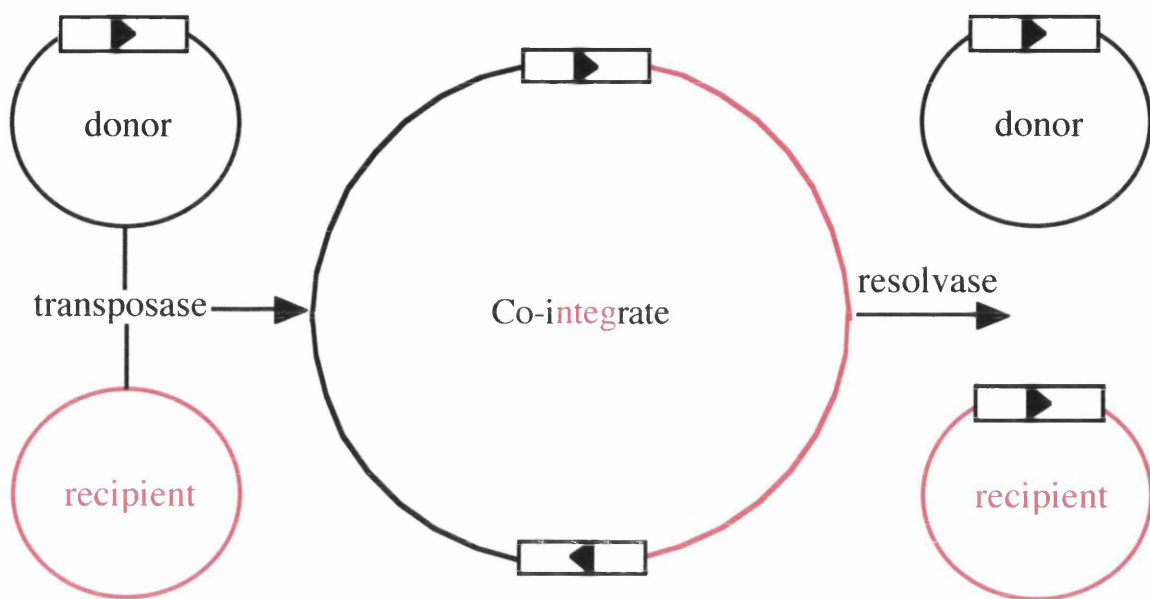
Tn3 is a bacterial transposon (illustrated in Figure 1.9.1) that carries all the components required for its survival and propagation. It is 4957 bp long and





**Figure 1.9 Transposon Tn3 and *res***

(1) Illustration of the Tn3 transposon. The arrows at the ends show the 38 bp sequence where the transposase binds (2) Diagram of *res*. The regions of the DNA strands that are broken during the reaction are shown in red. The sites where the resolvase binds are shown by black triangles at the ends of each site.



**Figure 1.10 The transposition of Tn3**

Illustration of the transposition of Tn3. The transposase proteins bring about the co-integrate structure that is resolved by the resolvase protein into two circles, each of which contains a copy of the transposon.

carries genes encoding the transposase enzyme (*tnpA*) and the resolvase enzyme (*tnpR*). Between these two genes is a sequence where the resolvase enzyme acts called *res*. The sites where the transposase acts are the 38 bp inverted repeats that flank the transposon. Tn3 also encodes a protein (Bla) that confers resistance to antibiotics that contain a  $\beta$ -lactam ring in *E.coli*. (Sherratt, 1989). Tn3 belongs to a group of  $\beta$ -lactamase encoding plasmids isolated from a range of enteric bacteria. It is likely that all these transposons have interchangeable functions, as they share more than 90% sequence similarity (reviewed in Sherratt, 1989).

Transposons of the Tn3 family move in two steps as illustrated in Figure 1.10. The first step requires the transposase (*tnpA* gene product) and involves fusion of the transposon-carrying molecule (the donor) with the site to where it is jumping (the recipient), to form a co-integrate structure. This co-integrate separates the donor and recipient molecules by directly repeated copies of the replicated transposon at each end (Gill *et al.*, 1979). The second step requires the resolvase (*tnpR* gene product) whereby it binds to the two *res* sites within each copy of the transposon and carries out a resolution reaction. The result of this is a segregation of the co-integrate into two molecules, each carrying a copy of the transposon.

In the case of  $\gamma\delta$  and Tn3, the *res* site is located between the divergently transcribed *tnpA* and *tnpR* genes, and the promoters for these genes are situated within *res*. In addition to its role in recombination, resolvase acts as a repressor of its own transcription and that of the transposase protein.

#### **1.14 The recombination site, *res***

*res* is a region of DNA 114 bp in length (see Figure 1.9.2). It contains three binding sites, I, II and III that each bind a dimer of resolvase. Each binding site is flanked by inverted repeats of an imperfectly conserved 9 bp sequence (consensus: T G T C Py Pu T T A). These binding sites are not evenly spaced, and their lengths differ because the spacing between the inverted repeats in each

binding site is not identical. The 9 bp inverted repeats span a central spacer of 10 bp in site I, 16 bp in site II and 7 bp in site III.

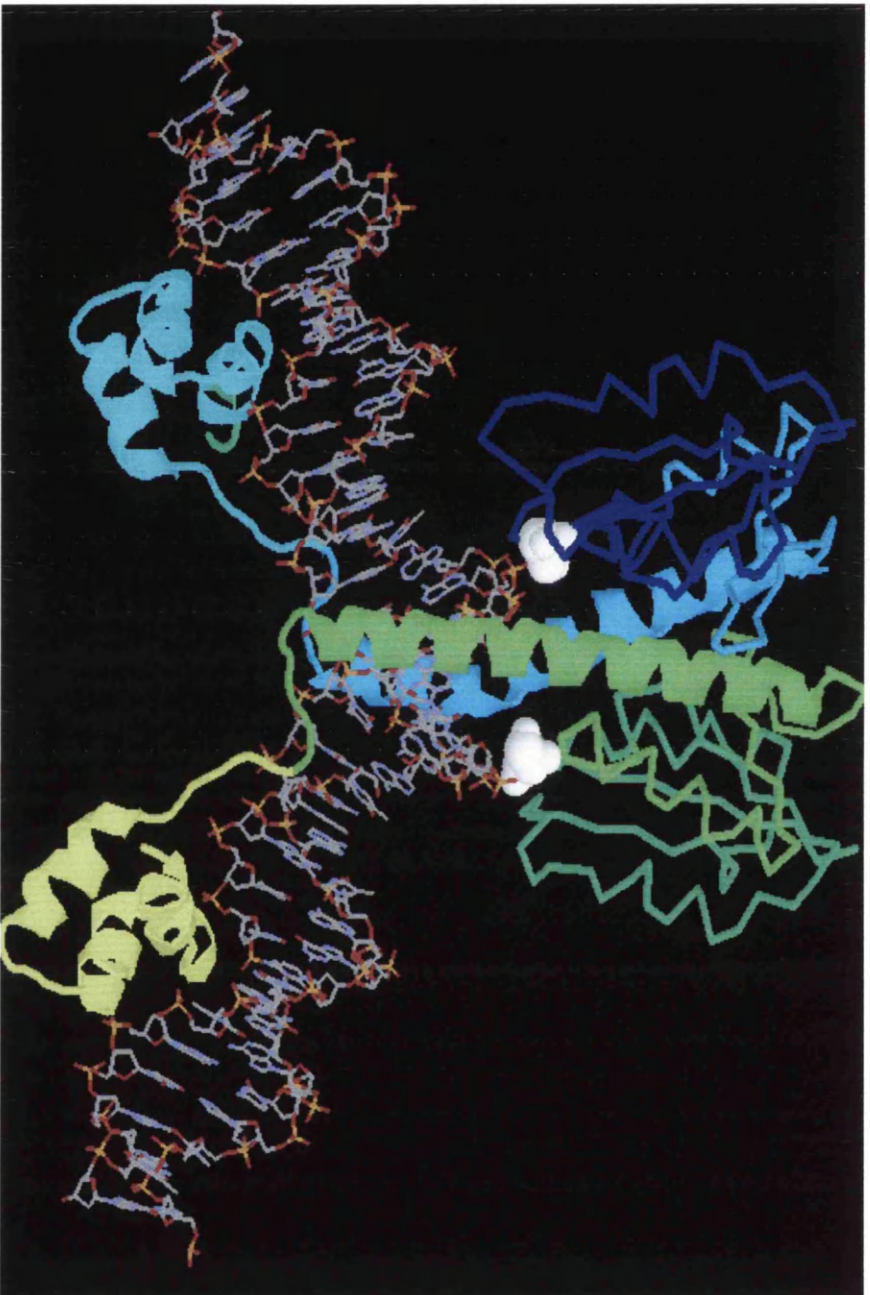
The structure of the *res* site is such that there are 53 bp or 5 helical turns between the centre of sites I and II, and around 3.3 turns of the helix between sites II and III. The importance of the helical phasing is such that the altering of the spacing between binding sites I and II can inhibit the recombination reaction, and most natural *res* sites have an integral number of helical turns between sites I and II (Hatfull and Grindley, 1988).

### **1.15 Tn3 resolvase**

Tn3 resolvase family is 185 amino acids long (20.5 kDa). It has two functional domains identified by chymotryptic digest (Abdel-Meguid, 1984) to be an N-terminal catalytic domain comprising residues 1-140 (15 kDa), and a C-terminal DNA-binding domain from residues 141-183 (5 kDa). An extended arm region at residues 121-147 connects the main parts of the two domains and binds the DNA in the minor groove (Yang and Steitz, 1995).

The crystal structure of the N-terminal chymotrypsin fragment (residues 1-140) has been solved at 2.3 Å (Sanderson *et al.*, 1990). The C-terminal domain was disordered 'a crystal' of the whole protein. The N-terminal domain contains the residues essential for catalysis and the residues involved in higher-order interactions between resolvase dimers (Hughes *et al.*, 1990 and 1993). It comprises a central 5-stranded  $\beta$ -pleated sheet surrounded by five  $\alpha$ -helices. Serine 10 is a catalytic residue situated in an unusual environment for an active site residue, as it is exposed on a protruding part of the protein structure, not buried in a pocket.

The C-terminal chymotryptic fragment contains a helix-turn-helix motif that binds in a specific way to the 12 bp consensus sequence in adjacent major and minor grooves, observed by footprinting studies (Abdel-Meguid *et al.*, 1984; Rimphanitchayakit *et al.*, 1989).



**Figure 1.11 Co-crystal structure of  $\gamma\delta$  resolvase with a 34 bp site I**

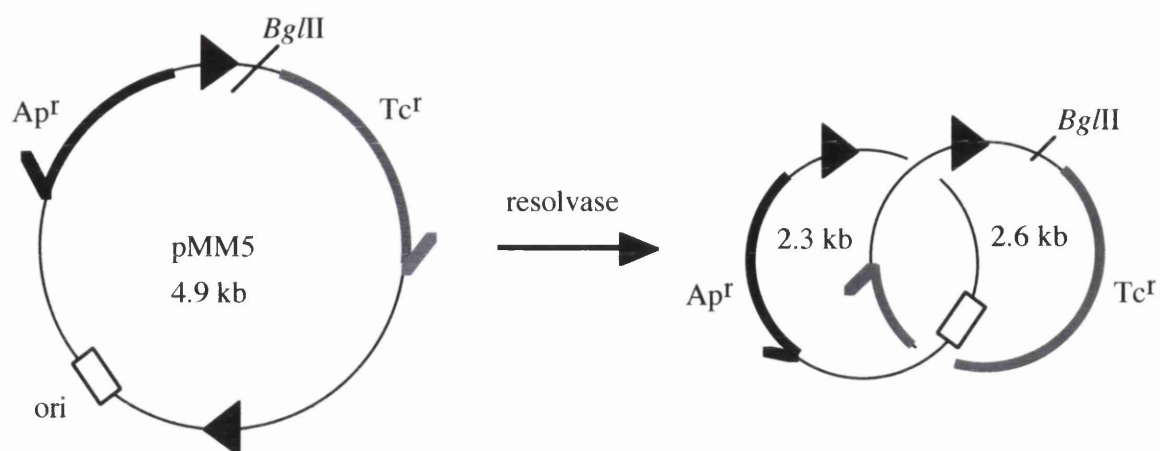
The resolvase monomers are shades of blue and shades of green, the nucleophilic serines (S10 and S210) are white. DNA is shown in wire frame (CPK). Residues 1-101 (201-301) are shown in backbone; 102-183 (302-383) secondary structure are shown in cartoon. Figure adapted from Yang and Steitz, 1995.

During the site-specific recombination reaction, resolvase cuts and religates the DNA within a region of perfect dyad symmetry at the centre of *res* sub-site I (Figure 9.1.2). This staggered double break in the centre of site I is the site where resolvase covalently attaches itself *via* a 5'-phosphoserine linkage (Reed and Moser, 1984). Binding sites II and III are known as the accessory sites which are responsible for bringing the two site Is into correct alignment during synapsis (Bednarz *et al.*, 1990).

### 1.16 The $\gamma\delta$ -resolvase-site I co-crystal structure

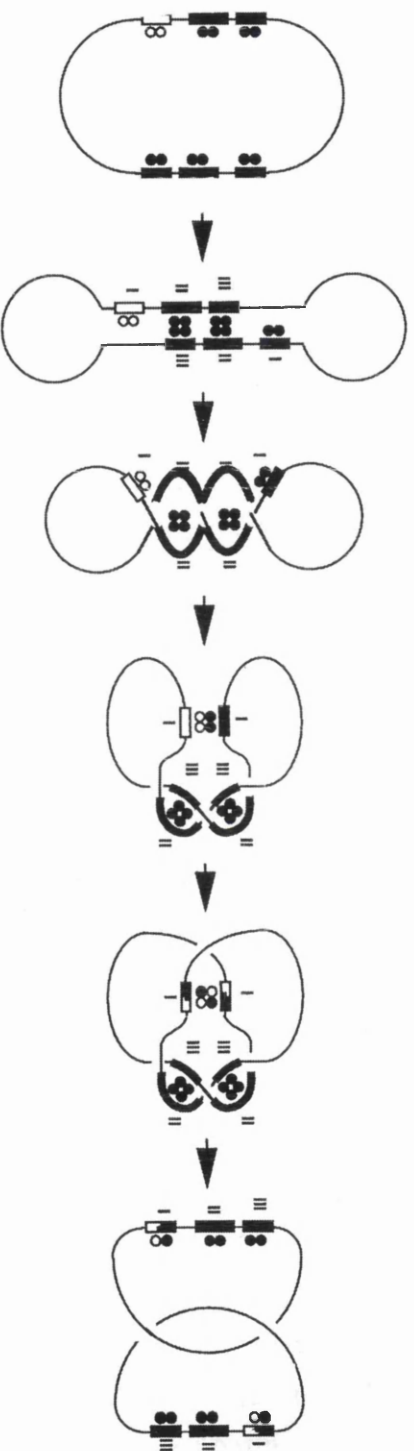
The co-crystal structure of  $\gamma\delta$  resolvase bound to a 34 bp oligonucleotides of site I is shown in Figure 1.11. Resolvase binds to site I as a 1, 2 dimer and induces a sharp bend of 60 ° at the centre of the DNA, towards the major groove. The catalytic residue was identified as serine at position 10 (Reed and Moser, 1984), and the equivalent serine at position 9 of the related recombinase Gin (Klippel *et al.*, 1988). The phosphates of the two DNA strands that become covalently linked to serine-10 lie approximately 13 Å apart across the minor groove of standard B-form DNA, whereas the  $\alpha$ -carbon atoms of the serine-10 in the 1, 2 dimer are 30 Å apart. From the co-crystal structure the hydroxyl groups of the serines are 17 and 11 Å away from the scissile phosphates. Thus a large conformational change is required to bring the serines close enough to the DNA before the phosphoserine bonds can be formed.

The region neighbouring serine-10 consists of conserved polar and positively charged amino acids, Tyr-6, Arg-8, Gln-14, Gln-19, Arg-68 and Arg-71,. Mutations in these residues abolished recombination (Hughes *et al.*, 1990), though reduced activity was present in some of the mutants. Therefore these residues are important for the formation of the active site or are directly involved in catalysis. In Tn3 resolvase there is a highly conserved tyrosine Y6 thought to be important for active site function. However the tyrosine hydroxyl group is not required for cleaving and rejoining (Leschziner *et al.*, 1995).



**Figure 1.12** The *in vitro* reaction catalysed by Tn3 resolvase

Illustration of the resolution of pMM5 by resolvase. The product is a -2 catenane.



**Figure 1.13 The topological filter model of synapsis**

*res* binding sites are shown as thick black lines (white in one subsite I only) along with their associated protein dimers. Random collision of the bound protomers occurs until the catalytic tetramers bound at subsites II and III come together in the antiparallel alignment. The DNA is then wrapped around these 'accessory sites' bringing about the site I-site I interaction. Any pre-synaptic complexes that do not have the correct topological requirements will not proceed to strand cleavage (Figure adapted from Stark and Boocock, 1995).



The main difficulty in relating the crystal structure of the site I-resolvase complex to the mechanism of strand cleavage is the position of the catalytic serines in relation to the phosphodiester bonds that are to be cut and covalently linked to the protein. The solution structure of the N-terminal domains indicated the presence of two kinds of crystallographic dimers, the 1,2 dimer and the 2, 3 dimer (Sanderson *et al.*, 1990). In the 1,2 dimer, which is thought to exist in solution (Hughes *et al.*, 1993), the catalytic serines are approximately 13 Å away from the phosphate backbone. The distance from the DNA is only known for the 1,2 dimer in the crystal structure bound to site I (Yang and Steitz, 1995). The catalytic serine residues in the 2,3 dimer are 17 and 11 Å from the DNA. One could envisage that during the course of the reaction the dimers bound at the crossover site change conformation from the 1,2 to the 2,3 dimer in order to facilitate strand cleavage *via* the catalytic serines. Or it could be that the bound resolvase dimers maintain the 1,2 dimer conformation, and gross conformational changes of the 1,2 dimer itself bring about strand cleavage. To date there is no biochemical evidence to support either theory; however the covalent fixing of subunits could perhaps make the problem clearer.

### **1.17 Binding properties of *Tn3* resolvase**

Site I of *res* is occupied by a dimer of  $\gamma\delta$  resolvase in the form of a 1, 2 dimer (Yang and Steitz, 1995). In  $\gamma\delta$  resolvase the protein binds to each binding site as dimers in a highly co-operative manner and the protein is thought to exist as a dimer in solution. In *Tn3* resolvase, binding is dimer-dependent and occurs in a co-operative manner (Blake *et al.*, 1995). The binding interaction occurs primarily through residues 153-183 which constitute the C-terminal DNA-binding domain. This region has been mapped through footprinting (Abdel-Meguid *et al.*, 1984) and ethylation interference analysis (Rimphanitchayakit *et al.*, 1989). The DNA-binding domain will be discussed further in Section 4.2.

### **1.18 *In vitro* reaction catalysed by Tn3 resolvase**

The *in vitro* system requires only a simple buffer, a supercoiled substrate containing two *res* sites in direct repeat and the resolvase enzyme (Reed and Grindley, 1981). DNA cleavage, strand exchange and re-ligation all occur at the centre of site I. Double strand cleavage results in a 2 bp overhang at the central AT nucleotide, resolvase being attached to the recessed 5' end by a phosphoserine bond (Reed and Moser, 1984). Once the strand exchange occurs, the phosphoserine bonds are broken to form the recombinant product. The product of the *in vitro* reaction is a (-2) catenane (Figure 1.12). The two circles are thought to be unlinked *in vivo* by a type II topoisomerase (Bliska *et al.*, 1991).

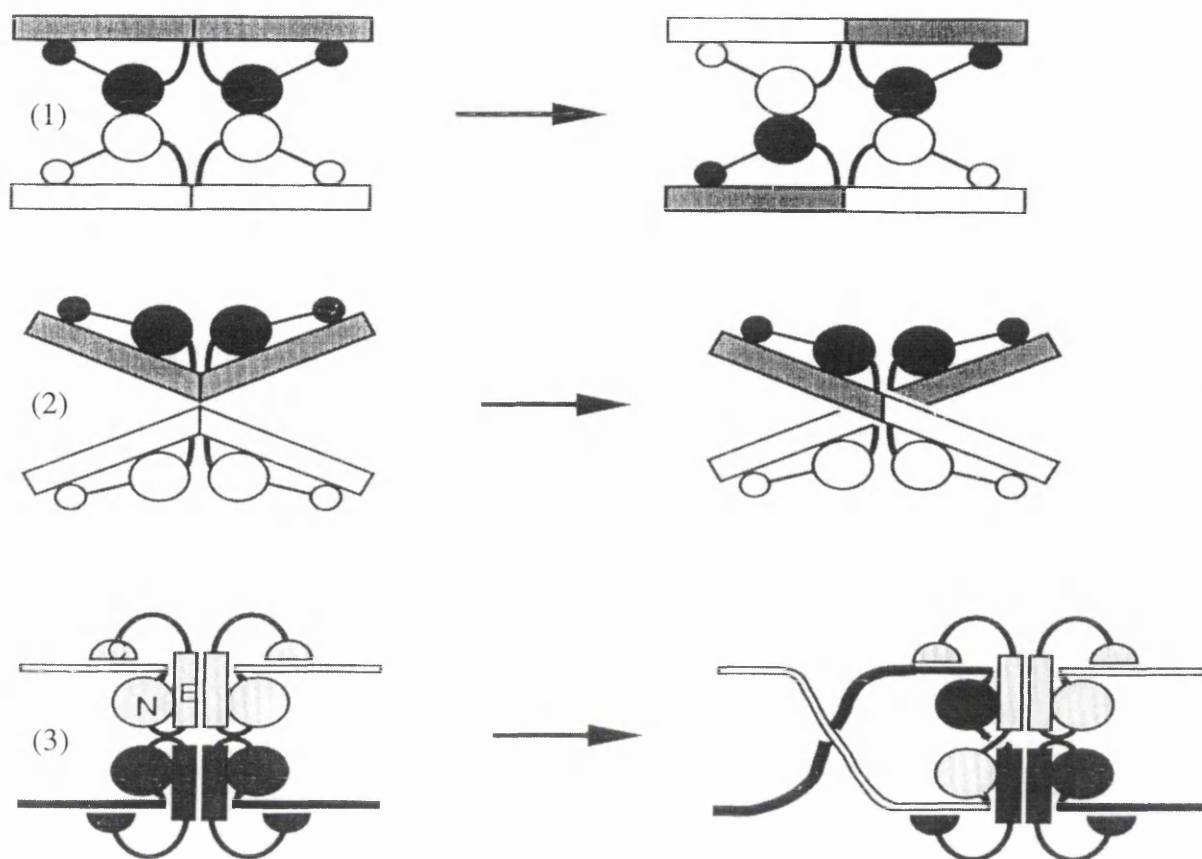
Three negative intra-domainal nodes are trapped between the *res* sites in the synaptic complex, and strand exchange is equivalent to 180° rotation in a right-handed sense (Figure 1.14). A linkage change of +4 occurs during resolution. The reverse is true of the reverse reaction (fusion) where the linkage change is -4 (Boocock *et al.*, 1987; Stark *et al.*, 1989; Stark *et al.*, 1991). The linkage change concurs with the sites being wrapped plectonomically around the protein, and strand exchange occurring by a simple rotation mechanism. Further topological analysis confirmed this result (Stark and Boocock, 1995).

### **1.19 Formation and functions of the synaptic complex**

Models for the way the synaptic complex is formed and for its structure have been proposed (reviewed in Stark and Boocock, 1995).

The topological filter or 'two-step synapsis' model explains the requirements for synapsis in terms of the possible productive and non-productive ways the DNA sites and the proteins bound to them can come together (Figure 1.13).

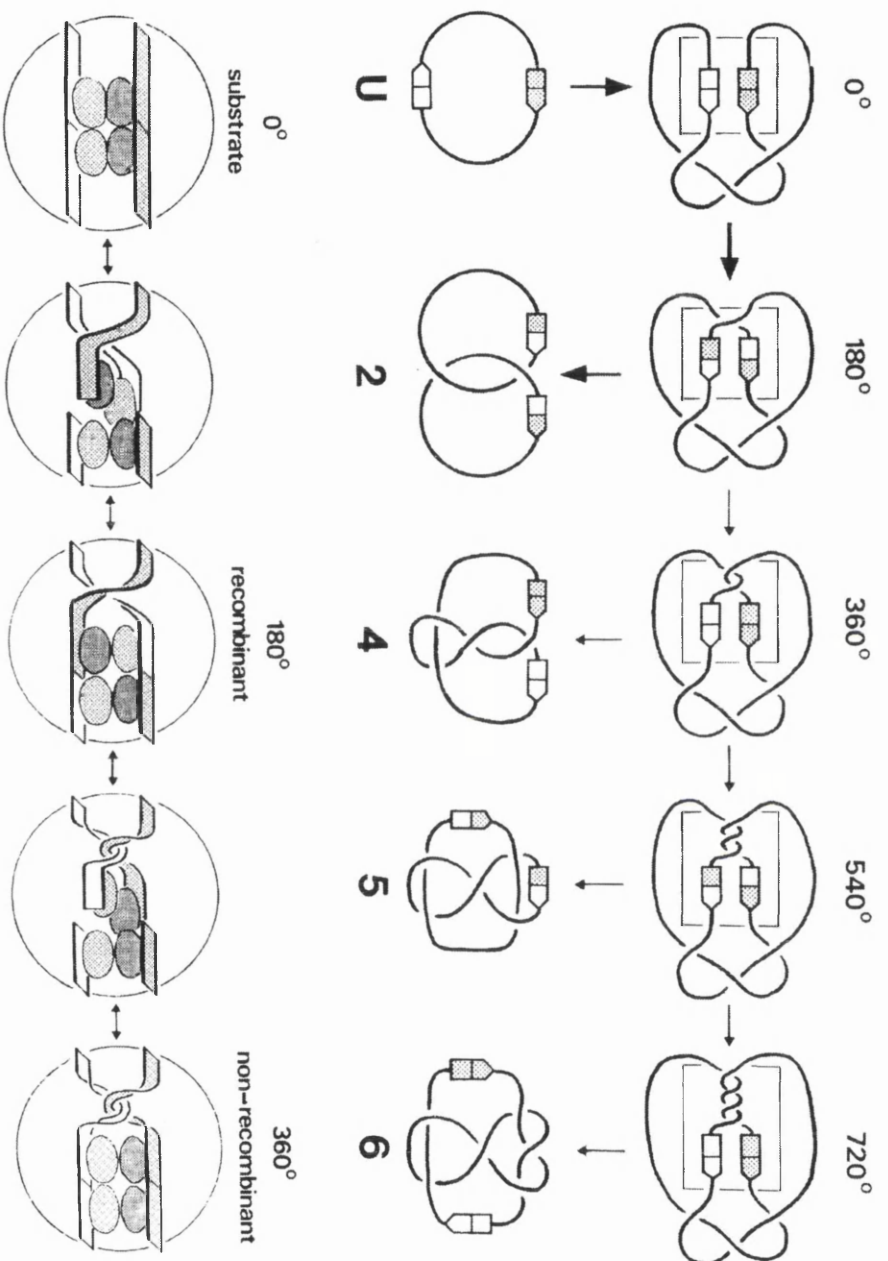
Here the *res* sites come together by random collision. However, for an active synapse to form which leads to recombination, the II and III accessory sites must



**Figure 1.14 Models of Strand Exchange**

- (1) subunit rotation
- (2) DNA-mediated
- (3) N-terminal domain rotation model

See text for details



**Figure 1.15 Multiple rounds of recombination**

Illustration of the catenanes and knots generated from multiple rounds of strand exchange. The amount of nodes is indicated by the numbers in bold (Figure adapted from Stark *et al.*, 1991).

intertwine in a defined way so that subsites I come into alignment, a process assisted by negative supercoiling. Incorrect synapsis structures involving the entanglement of DNA (such as from sites in inverted repeat orientation) would be energetically costly and would dissociate, allowing productive synapses to form. Analysis of synaptic structures of various substrates has provided support for this model (Watson *et al.*, 1996). Synapsis can occur with plasmids with sites inverted repeat orientation and with nicked substrates, although the overall yield of synaptic complex is much lower.

Resolvase dimers bound to all three subsites of *res* must come together, or synapse, before strand exchange can occur. Sites II and III form two tetramer cores that interwrap the DNA, trapping three negative nodes. This wrapping also brings the site I, with attached dimers, into alignment, a necessary accessory site function (Bednarz *et al.*, 1990). The site II/III component of the synaptic complex is thought to be involved in activation of the catalytic tetramer at site I. It is unclear if the activation is by protein-protein contacts between tetramers at II/III and I, or whether it is purely architectural, depending on the conformations of tetramers at II/III or I, relative to each other. Therefore the accessory sites, and the dimers bound to them, play an important role in the synaptic complex, being involved in topology, alignment and activation.

## **1.20 Models for Strand Exchange**

An illustration of the current models are shown in Figure 1.14.

### **Subunit rotation (Figure 1.14.1)**

The fact that the linkage change during resolution is +4 suggests a model in which the DNA is situated on the outside of the complex and the two resolvase monomers interact via their N-terminal domains on the inside. At strand exchange, two protein monomers rotate by 180° in a right-handed motion. A difficulty with subunit rotation is how the four cut DNA half-sites could be held together while rotation is occurring. This model would also require dissociation

of all four half sites. Experiments with mismatched substrates have shown that multiple rounds of recombination can occur, which fits well with this model. By using a substrate containing a mis-matched overlap in the 2 bp central dinucleotide, it has been shown that two rounds of recombination can take place without rejoining of the DNA strands in the intermediate recombinant form (McIlwraith *et al*, 1997). Indeed multiple rounds of recombination (Figure 1.15) can occur, which is explained very well by subunit rotation.

#### **DNA-mediated strand exchange** (Figure 1.14.2)

The two copies of the subsite I are brought into close proximity at the centre of a catalytic complex, with the bound protein dimers on the outside. This mechanism involves minimum movement of both protein and DNA during strand exchange. After resolvase-induced unwinding of the crossover site DNA, strand exchange is proposed to occur *via* movement only of the free 3' hydroxyls generated by cleavage, rather than by subunit exchange. It is difficult to imagine the four ends of DNA within the proteins being held together and coming together in the correct conformation. In addition, this theory does not explain iterative or multiple rounds of recombination since the DNA would become too tangled without a resetting mechanism.

#### **N-terminal domain rotation model** (Figure 1.14.3)

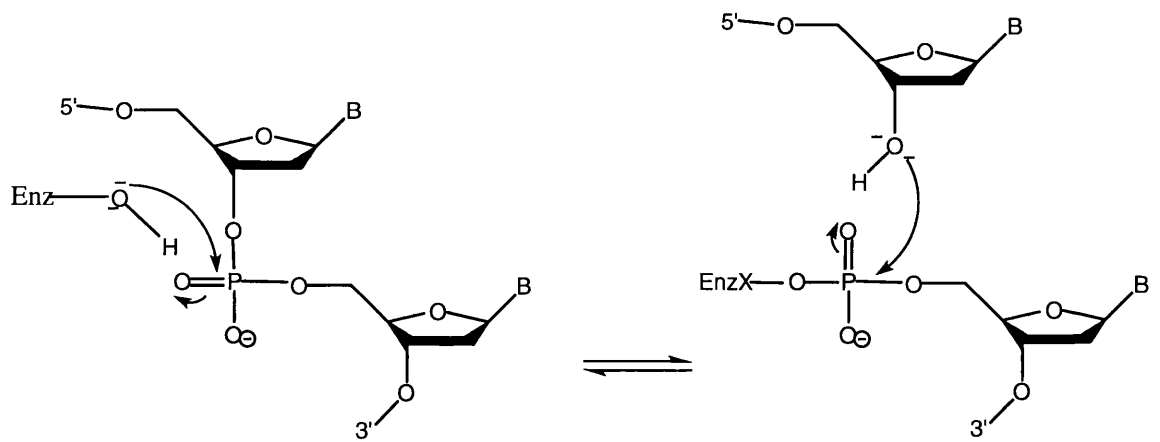
Mutants at resolvase positions 101-103 have been identified, that can resolve substrates lacking accessory sites II and III (Arnold, 1996). These "hyper-active" mutants can carry out resolution reactions on a plasmid containing two site Is only. This region might be a flexible hinge between the arm region (E-helix) and the N-terminal domain, allowing rotation of the N-terminal domains without dissociation of the 1, 2 dimer. Strand exchange would occur by rotation of the N-terminal catalytic domains by 180°. The arm region (residues 121-147) interactions remain intact. During domain rotation, the DNA-binding domain of one monomer in each dimer needs to dissociate to allow the DNA to escape. The

reaction can only proceed for two rounds of recombination without dissociation, unless there is a reset mechanism.

In view of the biochemical and topological evidence it is not clear which model is the true one. However the subunit rotation looks the most likely. The fact that it can account for the minor iterative products that appear in low amounts in addition to the two noded catenane shows that multiple rounds of rotation can occur without a re-ligated intermediate which would reset the dimers into a catalytic conformation. It has long been accepted that the strand exchange is equivalent to an 180 ° clockwise rotation, within the context of a synaptic complex that traps three negative nodes giving the unique major topological product of the -2 catenane. An expansion of this 180 ° rotation was to use mismatched substrates to induce two rounds of strand exchange yielding the predicted 4-noded knot. The observed 360 ° rotation, without a ligated intermediate (McIlwraith *et al*, 1997) could easily be explained by subunit rotation, whereas the DNA-mediated model would be very unlikely as the DNA would have a full turn of the helix between the dimers bound at site I, causing great structural anomalies. The catalytically-active conformation of the dimers after one round of N-terminal domain-mediated rotation is lost after one round of strand exchange, so it is unlikely that a second round would occur. Thus subunit rotation accounts nicely for the observed product topologies, but how the resolvase structure would allow subunit rotation to occur is still unclear.

### **1.21 The mechanism/chemistry of strand exchange**

Strand exchange (Figure 1.16) involves nucleophilic attack of Ser-10 residues on the two strands of the DNA duplex and two transesterification reactions, giving rise to a covalent resolvase-DNA complex, wherein the protein is attached to the 5'-recessed ends, giving rise to free 3' hydroxyls on each strand. Strand exchange then takes place and these 3'-hydroxyls attack the serine-DNA phosphodiester which results in the phosphate backbone rejoining, releasing the resolvase. In the reverse of the cleavage step (i.e. rejoining), the attacking hydroxyl group can be from the same DNA strand, giving a non-recombinant product, or from the



**Figure 1.16 The chemistry of resolvase-mediated strand exchange**

Illustration of the series of nucleophilic attacks that give rise to the covalent protein-DNA intermediate (Shown in one strand only).  
 Enz=the recombinase subunit; B=base and X=hydroxyl.



corresponding strand of the other cleaved site, giving a recombinant product. In Figure 1.16 only one strand of DNA is shown.

No Holliday junction intermediates have ever been isolated from Tn3 resolvase reaction, and it is thought that all four strands must be cleaved before strand exchange takes place. The exact order of cleavage events of all four strands is unclear at the moment. The popular theory is that all four strands are cleaved in a concerted array. However this type of concerted double-stranded cleavage would require a large conformational change of the catalytic dimer bound at the crossover site to bring the serine 10 nucleophiles closer to the scissile phosphates.

It was determined that the resolvase that is bound to each end of site I, i.e. site I left and site I right, is responsible for cleaving the half site to which it is bound (*cis* cleavage). This is in contrast to the Flp recombinase that cleaves the opposite strand to which the protein is bound (*trans* cleavage). It has also been shown that the top strand is preferentially cleaved before the bottom strand (Boocock *et al*, 1995), although all four strands must be cut before the strand exchange reaction can proceed. A difficulty with subunit rotation is how to imagine the mechanism whereby all four cleaved DNA ends are held together after cleavage and immediately prior to strand exchange. If the DNA strands are cleaved in sequence (i.e. top strand before bottom), then this may provide a means for concerted cleavage, so that only when the last strand was cut could rotation commence.

In light of the crystal structures and biochemical evidence to date, most has been learned about the mechanism of resolvase from the topological analysis of the products. A mechanism for the production of a -2 catenane was derived from a model whereby the least energetically costly synaptic structure was formed, and following a 180 ° clockwise rotation the -2 catenane was formed. The least energy requirements and minimal DNA entanglement were from a synaptic structure that trapped three negative nodes.

The topology of the synaptic structure was confirmed when experiments were carried out with *loxP* sites substituted for *res* site I. Addition of resolvase formed the expected synaptic structure at sites II and III trapping three negative nodes. Addition of the Cre invertase which cleaved at the *loxP* sites, gave the predicted topological products for an inversion reaction in the context of a synapse with three negative nodes (Kilbride *et al.*, 1999).

### **1.22 Crosslinking techniques to study binding/synapsis and strand exchange**

Using a combination of the data obtained from crystallographic, mutational and biochemical analysis of the recombinational apparatus, quite a lot of information can be deduced about the regions of the protein that come into close contact with the DNA during recombination. Residues that are important for catalysis have been identified through mutational analysis, and similarly residues essential for binding have been identified.

Site-specific crosslinking as a method of analysis of the recombination reaction is unique in that physical constraints can be imposed on the protein-DNA interaction that can allow us to "fix" the interacting components. The formation of covalent linkages between the protein and the DNA it associates with is a powerful technique for structural and functional studies. Photocrosslinking experiments (McIlwraith *et al.*, 1996) have shown that at least one of the four resolvase monomers bound at the crossover site can remain physically attached to the same DNA half site before and after recombination.

If high yields of crosslinking are obtained, a similar experiment to the above could be carried out involving the crosslinking of two protein monomers to binding site I. If the recombination reaction is inhibited by the covalent attachment of both ends of site I, it would suggest that protein dissociation from these sites is a requirement for recombination. Therefore the N-terminal domain rotation model would look more probable. However, if the recombination reaction worked with both sides covalently attached, then subunit rotation would

be validated, as a double round of strand exchange would only be consistent with subunit rotation (not with the DNA-mediated model).

If a reaction where a covalently attached mis-matched substrate was crosslinked at both ends of site I underwent strand exchange, this would also validate subunit rotation and that the presence of a re-ligated intermediate was not required.

Additional experiments could be performed that crosslink site II and III to targeted accessory site resolvases. Experiments analogous to those of Murley and Grindley, 1998 could be used to confirm current results about the requirements in the protomers bound to *res*, for all aspects of synapsis and recombination.

Site-specific crosslinking of components of the synaptic complex could provide useful information about the contacts that are made and indeed maintained during synapsis. Essential contacts could be identified by crosslinking the DNA at specific residues within the synaptic complex and direct inferences about the architecture of the components involved could be made.

### **1.23 Aims of research**

In order to investigate the mechanism of strand exchange, two different approaches were adopted. The first involved developing a methodology of crosslinking the protein to its binding site, by modifying the DNA-binding site at points where it is known to form close interactions with specific residues in the DNA-binding domain of the protein. Once a covalent DNA-protein binding complex is formed, the different modes of strand exchange could be tested by introducing other inter-subunit crosslinks at different positions. This approach was to be tested on resolvase, but, if successful, would be applicable to many other experimental systems involving proteins that interact with DNA.

The second approach involves introducing a chemical modification at the point where the resolvase cleaves. By placing a chiral phosphorus atom at the point where the strand breakage occurs, it should be possible to determine the number of catalytic steps by relating to the stereochemistry of the phosphorus atom before

and after cleavage and strand exchange. The work of this thesis is directed towards these aims, and the following steps will be described:

1. The chemical synthesis of thiol-containing phosphoramidite reagents.
2. Incorporation of these reagents into synthetic oligonucleotides that code for the binding sites of Tn3 resolvase.
3. Construction of expression vectors containing the DNA-binding domain of Tn3 resolvase as a removable cassette.
4. Expression and purification of novel cysteine mutants of Tn3 resolvase.
5. Crosslinking *via* forming disulfides between protein cysteine thiols and the thiol-containing oligonucleotides.
6. Synthesis of synthetic site I oligonucleotides with a phosphorothioate modification at the phosphorus atom that becomes covalently linked to the resolvase nucleophile during the cleavage of the DNA strands.
7. Incorporation of these modified site Is into supercoiled substrates and assaying them with resolvase protein.

## **Chapter 2: Synthesis of thiol-containing nucleotide analogues**

## 2.1 Introduction

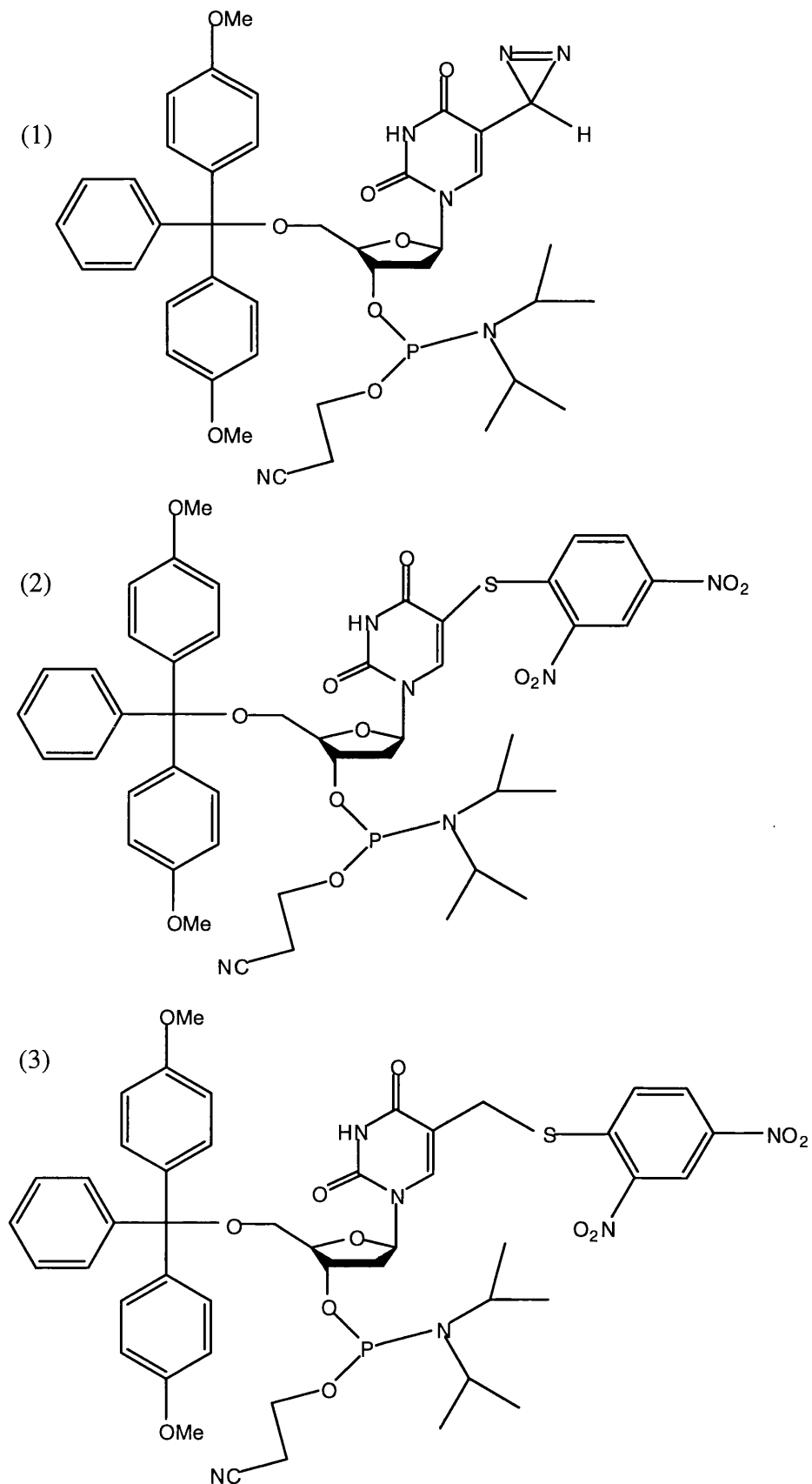
All of the nucleotide analogues that were to be used in crosslinking experiments carry the crosslinking group at the C-5 position of the pyrimidine ring. This position is not involved in the formation of Watson and Crick base pairs (Yuriev *et al.*, 1999). These compounds are structural analogues of the natural base thymine, which contains a methyl group at the C-5 position.

Two types of modification were to be made. One involves the incorporation of a photoactivatable group at the C-5 position, whereas the other introduces thiol moieties at C-5.

These analogues were to be incorporated into oligonucleotides and then activated to crosslink Tn3 resolvase (Chapter 5). The photochemical target (Figure 2.1.1) can be activated with long wave UV light; the thiol-containing nucleotides are designed to crosslink *via* the formation of disulfide bonds. The latter strategy required the creation of novel cysteine mutants of Tn3 resolvase (Chapter 4), in regions where it is known to form close interactions with the DNA.

There are potentially two different ways to make nucleotide analogues. The first is to modify the nucleotide directly. The second is to modify the base, then couple it to the sugar portion (Wilson *et al.*, 1995). These coupling reactions usually give two anomeric nucleotide products, the  $\alpha$  and  $\beta$  forms. Through modification and optimisation of the reaction conditions it is usually possible to favour one anomer over the other, although their separation is often difficult. It was preferred to use the nucleotides that already contain the desired anomeric conformation. The starting materials for the synthetic targets (Figure 2.1) were thymidine and 2'-deoxyuridine.

For crosslinking, the nucleotide analogues must be incorporated into oligonucleotides (Chapter 4). The phosphoramidite analogues can be incorporated into the oligonucleotides site-specifically. The phosphoramidite-protected thiols



**Figure 2.1 Illustration of synthetic targets**

The three phosphoramidite targets. (1) is photoactivatable, whereas (2) and (3), once deprotected and incorporated into oligonucleotides, should form disulfides with engineered cysteine residues in the protein.

and diazirine analogues shown in Figure 2.1 were to be made and tested as crosslinking agents.

The diazirine-modified nucleotide has the advantage of being activated by UV light at wavelengths between 350-380 nm. This is well above the wavelength at which the groups in proteins and DNA are activated (260 nm). Diazirines, upon UV irradiation form a highly reactive carbene radical (with the release of N<sub>2</sub>), that can form stable adducts with most amino acids groups in proteins (Lui, 1986), so it should be easy to induce crosslinking at any chosen position. The diazirine groups are quite small, therefore should cause minimal unfavourable structural changes within the protein/DNA complex.

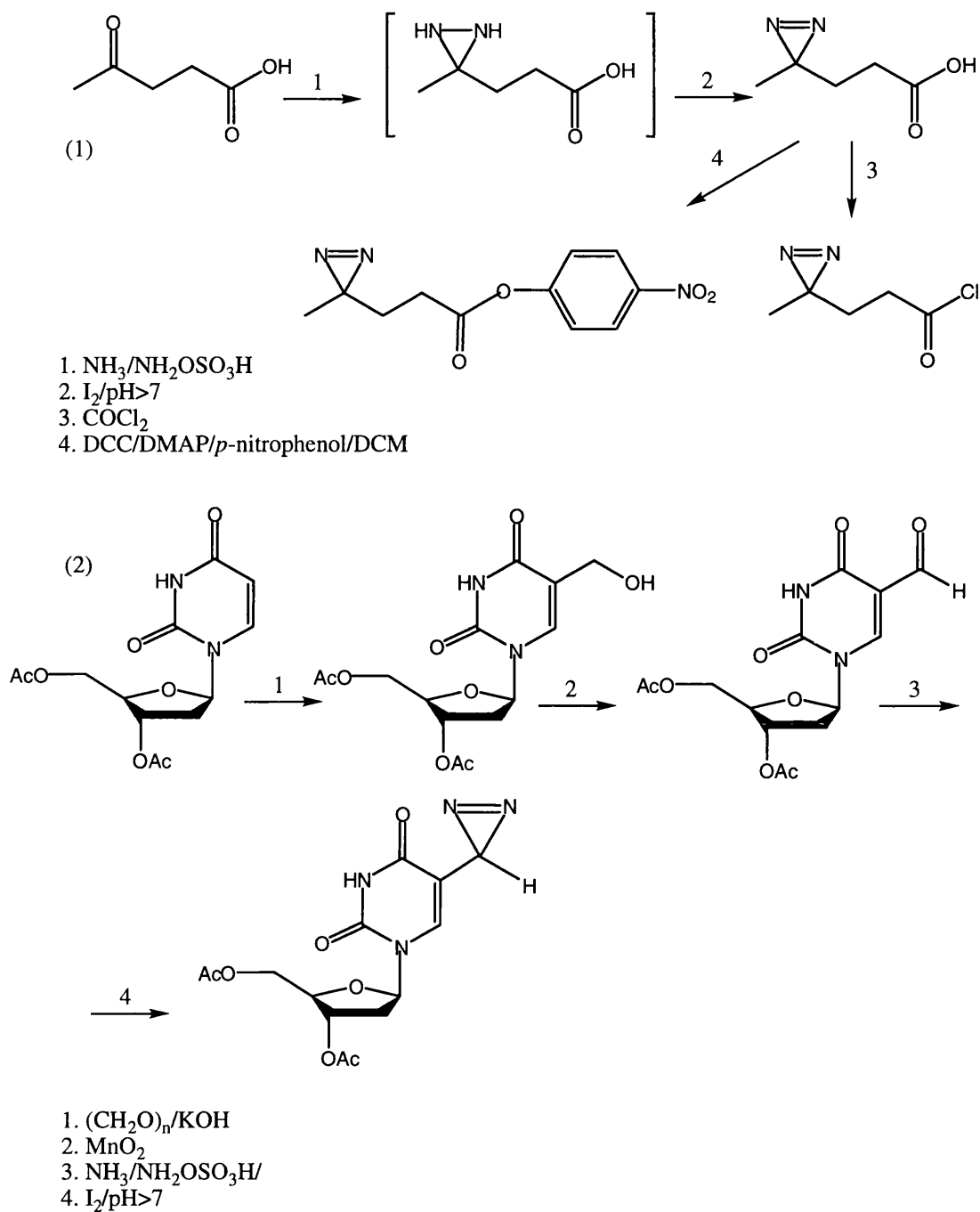
Thiols are potentially excellent site-specific crosslinking agents. The thiol-disulfide interchange proceeds readily under physiological conditions, and is reversible and selective. Two reactant thiols can be encouraged to form disulfides using a variety of oxidising agents. Thiols in the DNA substrates can be positioned close to engineered cysteine residues in proteins (Chapter 4).

The thiols require suitable protecting groups, as they are very nucleophilic and hence reactive. The protecting group must be stable enough to survive the conditions of the DNA synthesis, and must be stable to oxidising/reducing agents. It should also be relatively easy to remove once it is incorporated into the oligonucleotides. The protecting group chosen for the thiols was 2,4-dinitrophenyl (Meyer and Hanna, 1996), as it fulfils the above criteria. The effect of the structure and stability of DNA hybrids containing the analogue 5-[S-(2,4-dinitrophenyl-thio)]-2'deoxyuridine have been studied (Yuriev *et al.*, 1999).

## **2.2 Synthesis of the diazirine-containing nucleotide analogue**

Diazirines can be synthesised from ketones in a two-step process (Church and Weiss, 1970; Figure 2.2.1). Treatment of the ketone with ammonia and hydroxylamine-*O*-sulfonic acid yields the diaziridine which can be oxidised to





**Figure 2.2 Illustration of the diazirine reactions**

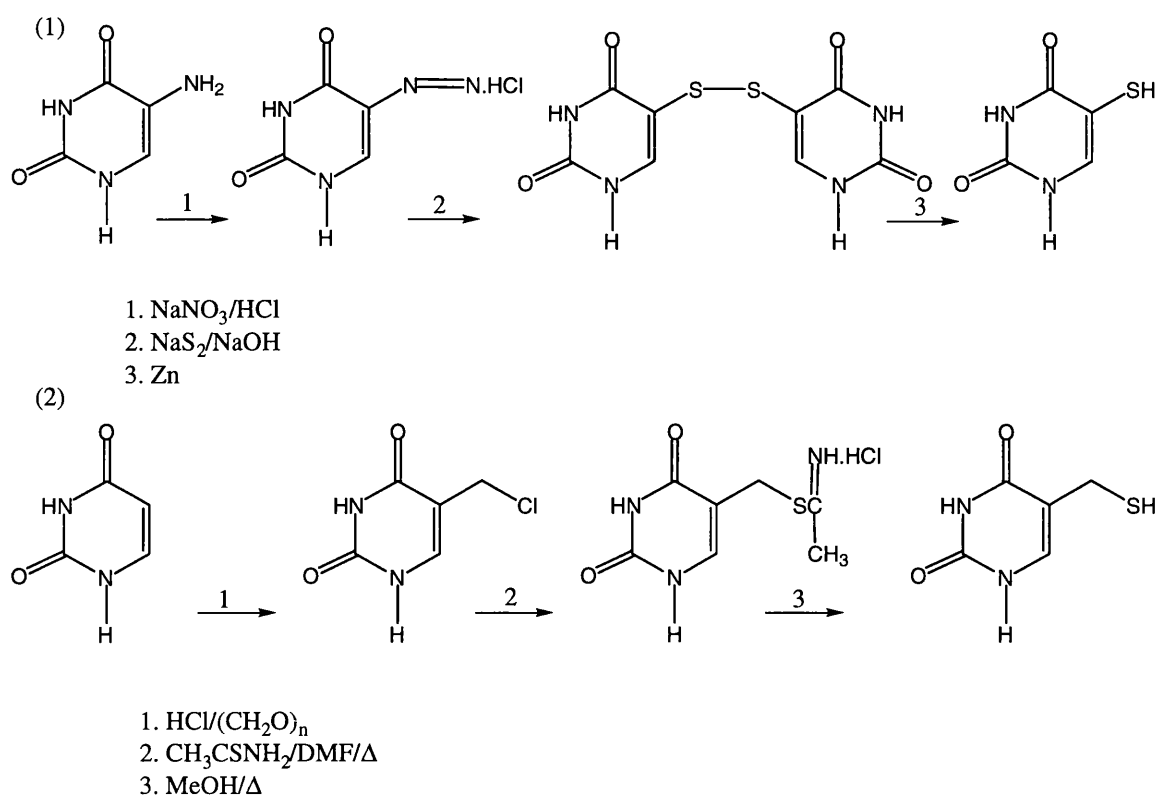
(1) Synthesis of the diazirine of levulinic acid and (2) proposed diazirine of 2'- deoxyuridine.

the diazirine using iodine and triethylamine. To make the diazirine target, it was necessary to first make the 5-acetyl derivative of 2'-deoxyuridine (Figure 2.2.2). An important precursor to the keto-compound is the hydroxymethyl compound, obtained by treating 2'-deoxyuridine with paraformaldehyde in the presence of base (Shiau *et al.*, 1980), by formaldehyde addition. The 5-hydroxymethyl-2'-deoxyuridine was oxidised to the 5-formyluracil nucleoside using activated manganese dioxide (Matulic-Adamic and Watanabe, 1988).

### 2.3 Thiol-containing pyrimidine compounds

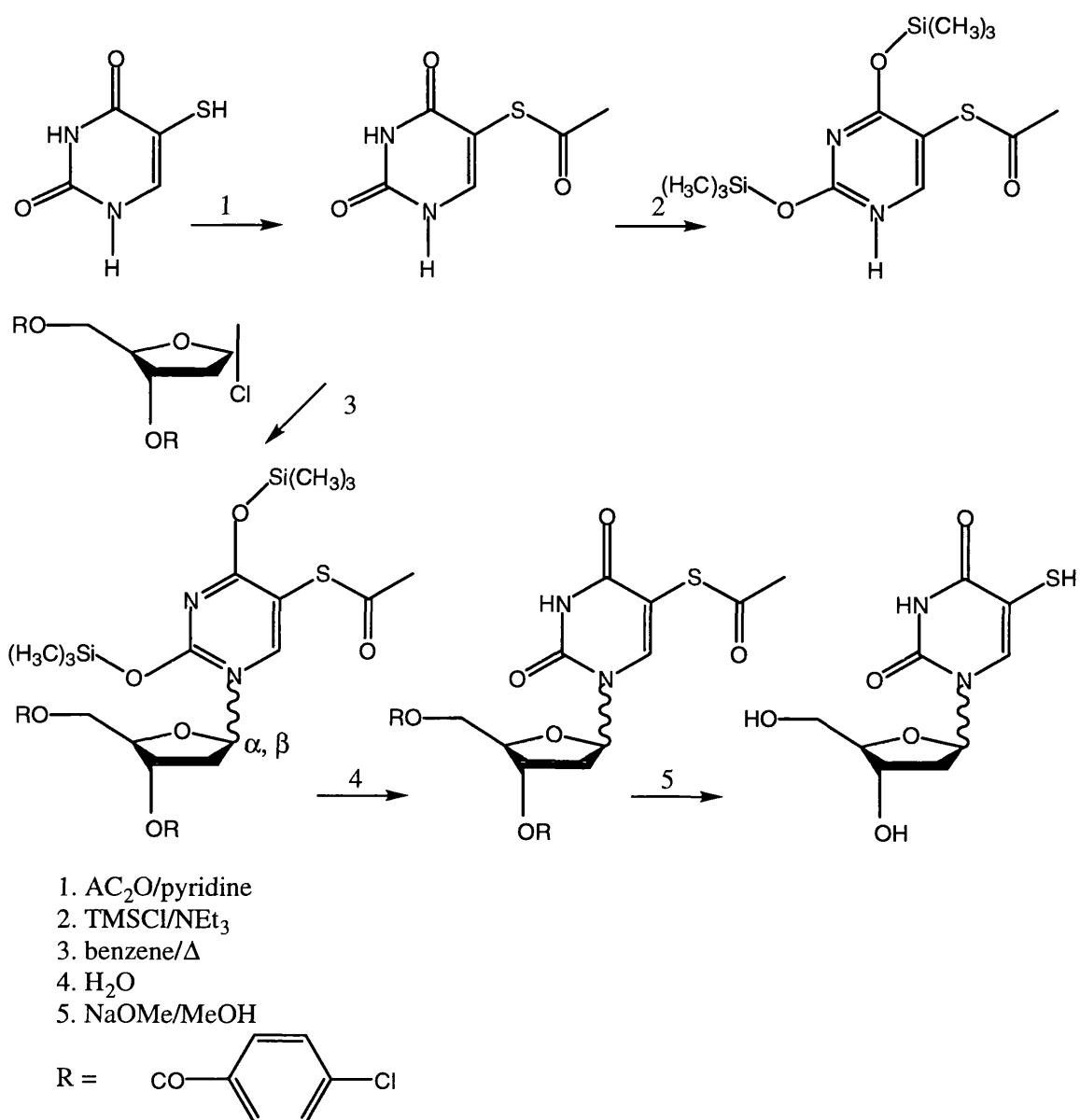
Reactions incorporating sulfur at the C-5 position of the pyrimidine ring were initially carried out on uracils. 5-Mercaptouracil and other S-substituted derivatives were synthesised (Bardos *et al.*, 1955; Figure 2.3.1). 5-Aminouracil was diazotised and using an alkaline solution of sodium disulfide, the 5-uracilyl disulfide was produced. This was reduced to the free mercaptan using zinc reduction. 5-Mercaptomethyluracil was also synthesised (Giner-Sorolla and Medrek, 1966 and 1978; Figure 2.3.2). The 5-hydroxymethyl compound of uracil was first halogenated, then reacted with thioacetamide in dimethyl formamide, yielding 5-acetiminomethyluracil hydrochloride. This was heated in methanol yielding 5-mercaptomethyl uracil.

Coupling reactions of thiol-containing bases (and other modified bases) to the sugar group are reviewed in Wilson *et al.*, 1995; Figure 2.4. The coupling of 5-mercaptouracil bis-*O*-trimethylsilyl derivatives to halogenated ribose compounds yielded anomeric 5-mercapto-2'-deoxyuridines (Bardos *et al.*, 1966). The desired  $\beta$ -anomer was isolated in 37.5% yield, although yields in excess of 70% (Dinan *et al.*, 1982) were also reported. The same methodology was also used to synthesise 5-mercaptomethyl 2'-deoxyuridine (Gupta *et al.*, 1975), and the desired  $\beta$ -anomer was obtained in 13-15% yield. Due to the difficulty in separating the anomers in these reactions, and the low yields of products, these reactions were omitted from this study.



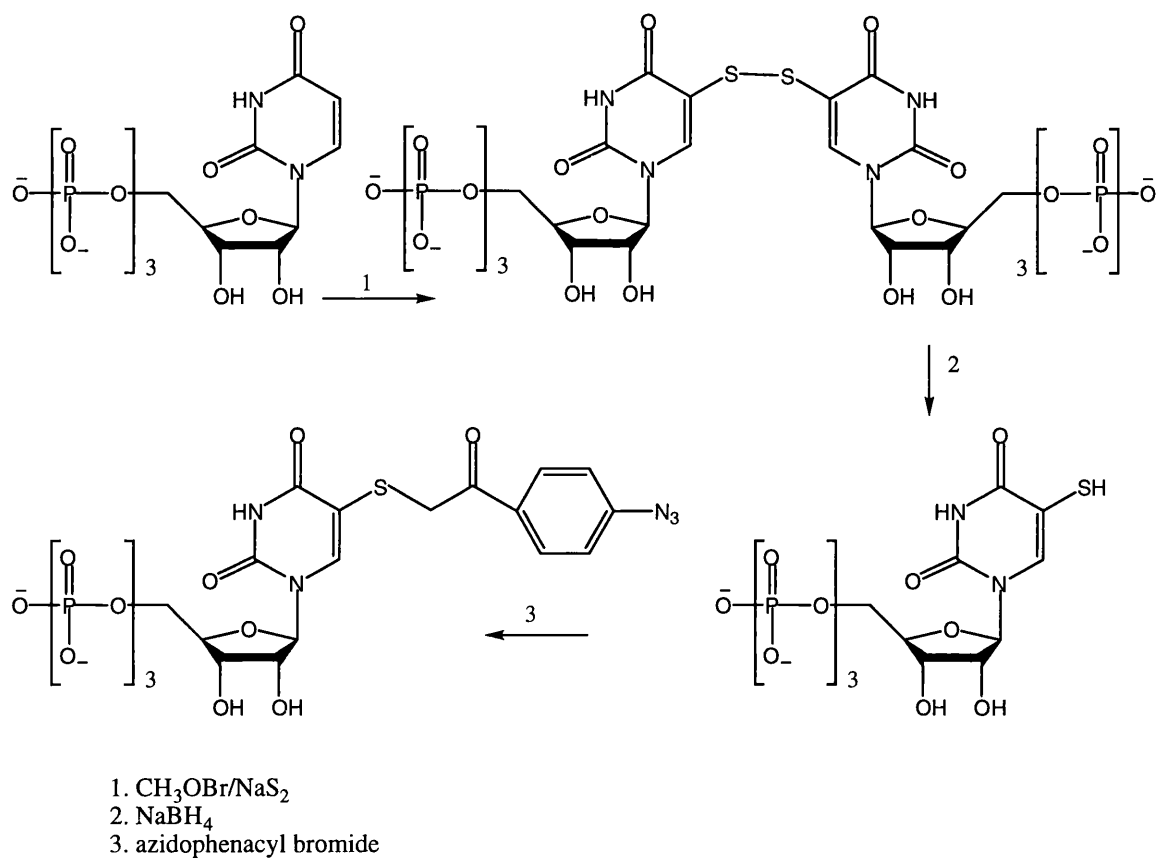
**Figure 2.3 Reactions yielding sulfur-containing uracils**

Scheme (1) shows the synthesis of 5-mercaptouracil from 5-aminouracil (Bardos *et al.*, 1955) and (2) illustrates the synthesis of 5-mercaptomethyluracil from uracil (Giner-Sorolla and Medrek, 1966).



**Figure 2.4 Coupling of sulfur-containing pyrimidines using glycosylation**

Illustration of the synthesis of the 5-mercapto-2'-deoxyuridine derivative (Bardos *et al.*, 1966). The equivalent coupling was carried out on the 5-mercaptomethyl derivative (Gupta *et al.*, 1975).



**Figure 2.5 Illustration of the synthesis of 5-[(4-azidophenacyl)thio]uridine 5'-triphosphate**

Diagram showing the synthesis of this photoactivatable nucleotide analogue (Hanna *et al.*, 1989).

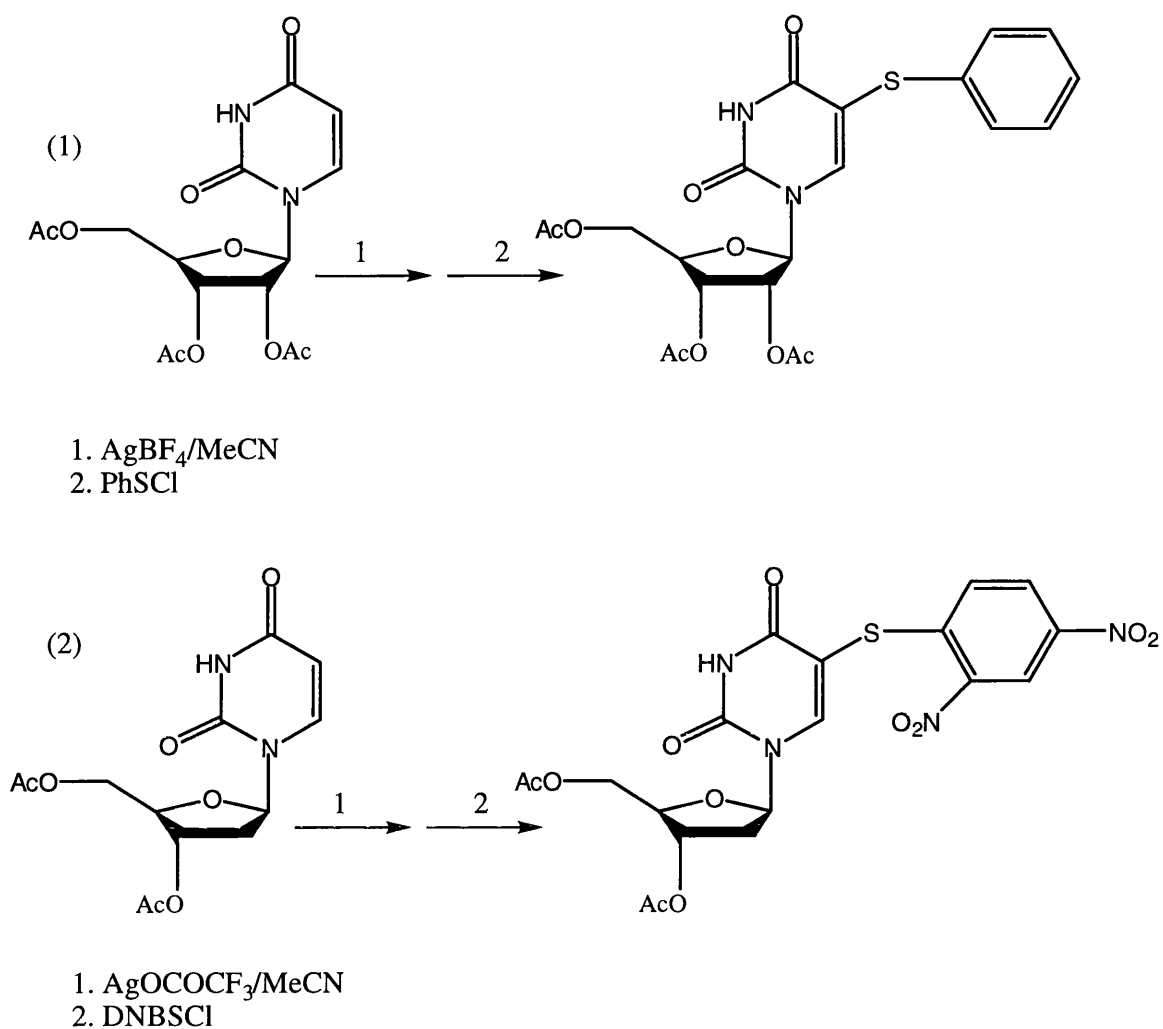
The nucleotide analogue 5-[(4-azidophenacyl)thio]uridine 5'-triphosphate was synthesised (Hanna *et al.*, 1989; Figure 2.5) using a different strategy. The 5-mercapto-UTP (5-SH-UTP) was synthesised by treating the triphosphate with methylhypobromide and sodium disulfide, resulting in the disulfide being formed. Following treatment of the disulfide with sodium borohydride, the thiol was obtained, which was subsequently alkylated with azidophenacyl bromide. The same methodology was used to synthesise 5-[(4-azidophenacyl)thio]cytidine 5'-triphosphate (Hanna *et al.*, 1993).

## 2.4 Direct Thioarylation Methods

These methods allow for the direct attachment of functional groups to the C-5 position of pyrimidine nucleotides. These syntheses are potentially useful, as they would shorten the phosphoramidite route by allowing direct introduction of the protected sulfur moiety on the pyrimidine ring.

For these syntheses it was necessary to synthesise *bis*-(2,4-dinitrophenyl) disulfide. Reacting 2,4-dinitrobenzenesulphenyl chloride with acetic acid, followed by aqueous work up resulted in the disulfide (Kharasch *et al.*, 1955). The disulfide can be reduced to 2,4-dinitrothiophenol by mild reduction with glucose and base (Bordwell and Andersen, 1953). This monomer unit would be useful as it could react with a halide and be used in the synthesis of target B, as a way of introducing the protected thiol group directly on the pyrimidine ring. The thiol reagent 2,4-dinitrobenzene sulfenyl chloride can be used as a source of the protecting group in these methods.

The use of silver reagents to catalyse the addition of arylthiols to the C-5 position was reported (Lee and Kim, 1991; Figure 2.6.1). This thioarylation reaction has been used on a range of uridines and 2'-deoxyuridines. The reaction of triacetyl-protected uridine with AgOCOCF<sub>3</sub> and benzene sulfenyl chloride resulted in C-5 arylation at 98% yield (Lee and Kim, 1991). The analogous reaction with 3',5'-*O*-diacetyl-2'-deoxyuridine will be used as a method of direct incorporation of the



**Figure 2.6 5-Thioarylation using silver reagents**

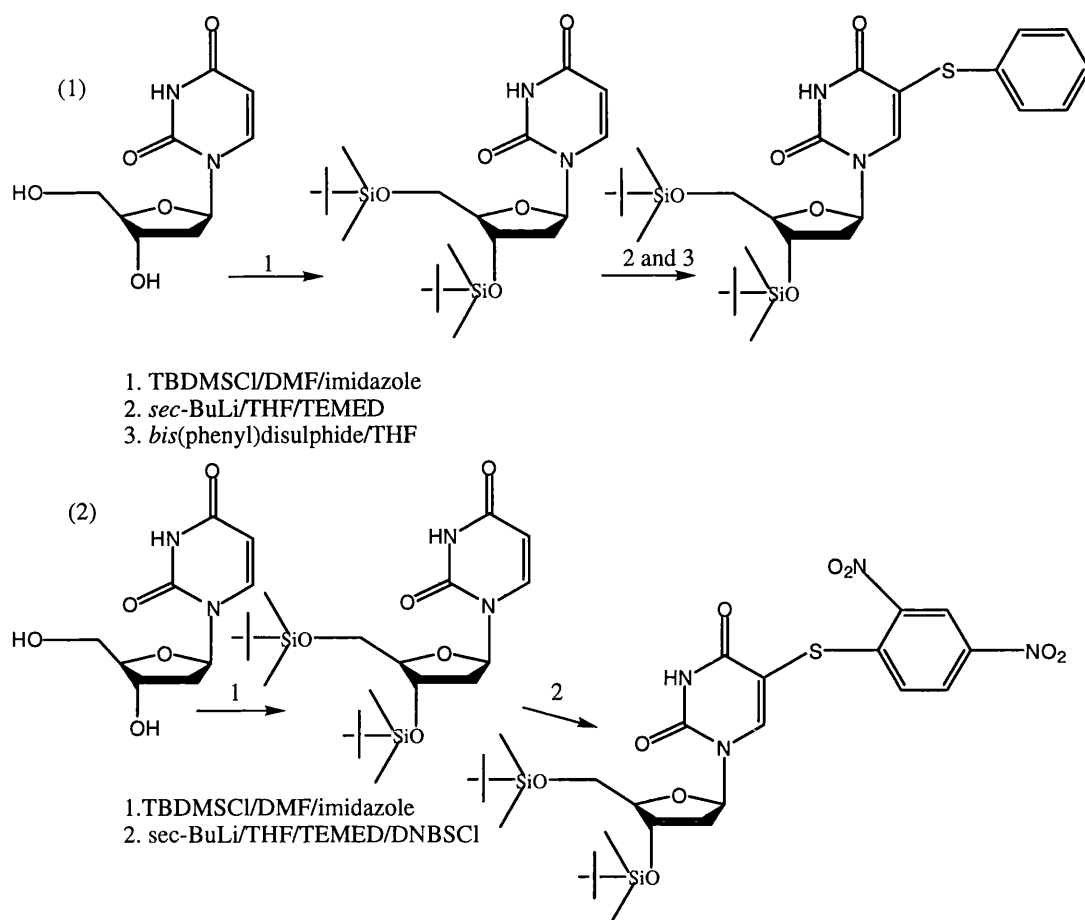
In (1) above the triacetyl derivative is thioarylated as shown and (2) is the proposed reaction for the diacetyl derivative.

protected thiol moiety (Figure 2.6.2). Higher yields of product were reported when  $\text{AgOCOCF}_3$  was used instead of  $\text{AgBF}_4$  for the 3',5'-*O*-diacetyl-2'-deoxyuridine derivative.

Another way of introducing functionality at the C-5 position of the pyrimidine ring is to make the lithium salt of the base, then quench it with different electrophiles. This reaction allows the electrophiles to be added in a regioselective way. Very high selectivity of C-5 over C-6 incorporation was observed when bulky lithium bases and *t*-butyldimethylsilyl protecting groups on the sugar hydroxyls of 2'-deoxyuridine were used (Ogilvie *et al.*, 1978; Armstrong *et al.*, 1989; Figure 2.7.1). The lithiation reaction on 3',5'-bis-*O*-(*t*-butyldimethylsilyl)-2'-deoxyuridine was quenched with different disulfides and electrophiles. When *bis*(phenyl)disulfide was used to quench the lithium salt, a 58% yield was found (Armstrong *et al.*, 1989). A  $^{13}\text{C}$ -methyl labelled thymidine was synthesised from 2'-deoxyuridine (Ahmadian and Bergstrom, 1998). The lithiation reaction was used in this work as it reduced the synthesis of the phosphoramidite targets by three steps. The lithium salt was quenched with either *bis*-(2,4-dinitrophenyl)disulfide or 2,4-dinitrobenzenesulfonyl chloride (Figure 2.7.2).

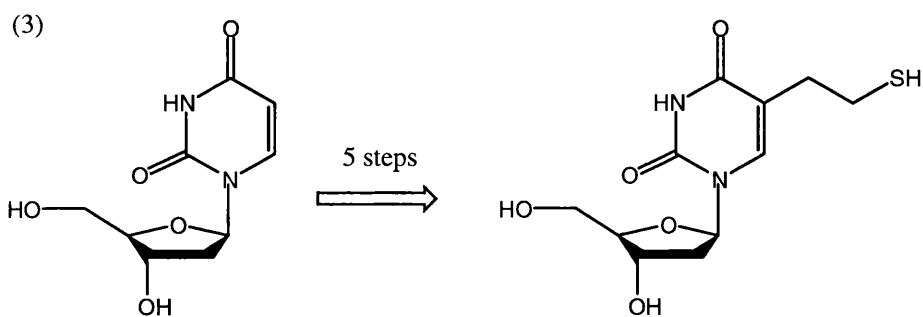
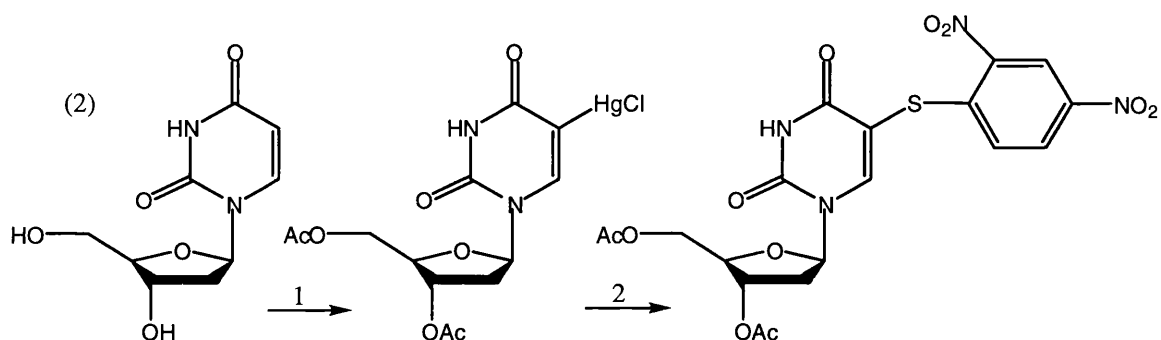
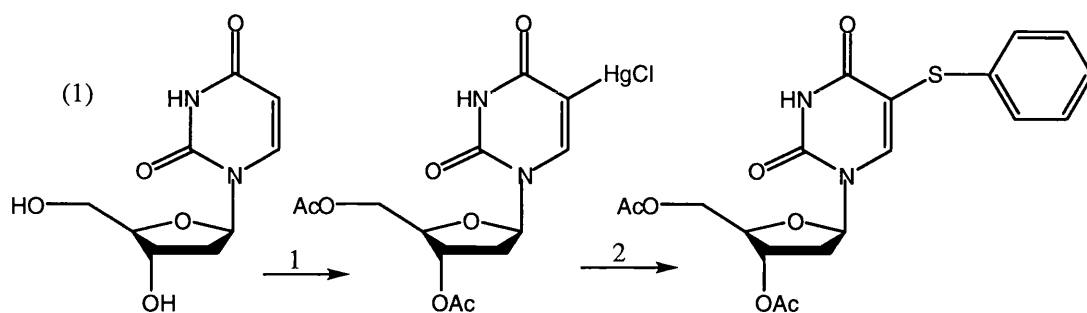
Another way of introducing functionality at the C-5 position of pyrimidines is by palladium cross-coupling reactions, using mercurated nucleosides with organic disulfides, acid and allyl chlorides (Bergstrom *et al.*, 1991; Figure 2.8.1). When aromatic disulfides were used, i.e. *bis*(phenyl)disulfide, the incorporation at C-5 was 52%. This reaction could potentially be used for introducing the protected thiol group in one step (Figure 2.8.2). A four atom linker was introduced using the coupling of allyl chloride using palladium, followed by oxidation, followed by thiol synthesis (Goodwin and Click, 1993; Figure 2.8.3). The starting material in these reactions is a very toxic mercuric salt, and the *bis*-(2,4-dinitrophenyl)disulfide is not soluble in the solvents required for the reaction, so this reaction was omitted from this study.





**Figure 2.7 Illustration of lithiation reactions**

Diagram showing the lithiation of protected deoxyuridine and quenching with various electrophiles. (1) shows quenching with disulfide reagents, and (2) illustrates the proposed reaction where the sulfenyl chloride is the electrophilic species.



**Figure 2.8 Palladium catalysed thioarylation**

In (1) the thioaryl group is added to the pyrimidine ring (Chang and Mertes, 1987; Bergstrom *et al.*, 1991); (2) shows the proposed reaction with *bis*(2,4-dinitrophenyl)disulfide and (3) shows the synthesis of a three atom alkylthiol-tether (Goodwin and Glick, 1993).

## 2.5 Indirect methods of thiol incorporation

Introduction of thiol moieties at the C-5 position of pyrimidine rings can be achieved *via* thiocyanation. A thiocyanation reaction with 2'-deoxyuridine yielded the 5-thiocyanato derivative (Nagamachi *et al.*, 1972; Torrence *et al.*, 1978). The analogous reaction with the 5', 3'-diacetyl-2'-deoxyuridine (Nagamachi *et al.*, 1974, Meyer and Hanna, 1996) resulted in much higher yields of the product. The reaction is an electrophilic substitution reaction of the pyrimidine ring using thiocyanogen chloride (Angus and Bacon, 1958). The electrophilic reagent is generated *in situ* using potassium thiocyanate and chlorine gas in anhydrous glacial acetic acid. An important factor in the reaction is that a stoichiometric amount of thiocyanogen chloride is reacted with the pyrimidine ring. This can be determined by titration of the chlorine solution with sodium thiosulfate solution (Flascha *et al.*, 1969). The reaction has also been carried out using lead thiocyanate as the source of thiocyanate (Bradley and Hanna, 1992).

For the synthesis of the 5-mercaptopomethyl-2'-deoxyuridine analogue, it was necessary to try a number of different methods (see Section 2.9). The 5-thiocyanato, thioacetyl and mercaptopomethyl derivatives of 2'-deoxyuridine have been made (Barwolff and Langen, 1975). These were made from the 5-bromomethyl precursor (Shiau *et al.*, 1980, Sergiev *et al.*, 1997), from reactions of the desired nucleophiles by simple S<sub>N</sub>2 reactions in dimethyl formamide.

## 2.6 Phosphoramidite Synthesis

For use in the DNA synthesiser, nucleotide analogues require specific protecting groups on the sugar hydroxyl groups. The 5'-primary hydroxyl on the sugar must be protected as a 4,4'-dimethoxytrityl (DMT) ether. These DMT groups are acid labile and thus easily removed during the DNA synthesis cycle, thus releasing the 5'-OH for the coupling step. 5-[S-(2,4-Dinitrophenyl)thio]-2'-deoxyuridine is treated with 4,4'-dimethoxytrityl chloride in pyridine (Bradley and Hanna, 1992). An increase in efficiency of incorporation of the DMT group was reported

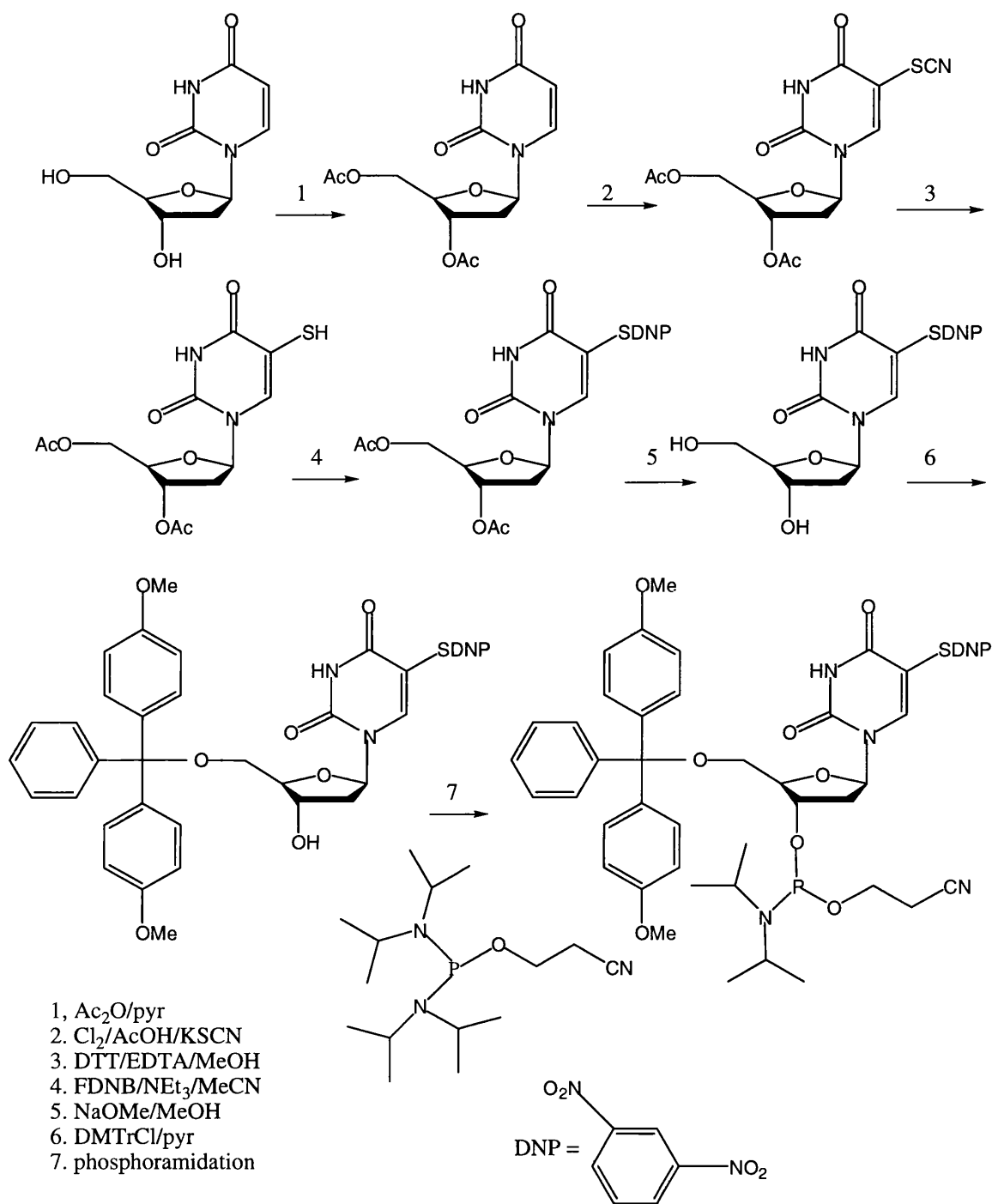
(Bleasdale *et al.*, 1990) when the trityl tetrafluoroborate salt was used for the protection step.

There are a number of methods for carrying out phosphitylation of the 3'-hydroxyl of pyrimidines. These use different phosphorus reagents and different activators. The activated phosphoramidite, 2-cyanoethyl-*N, N'*-diisopropylchloro phosphine (Sinha *et al.*, 1983) when reacted with the alcohol in the presence of diisopropylethylamine, yields the phosphoramidite. Another source of the phosphoramidite is the dimer, 2-cyanoethyl-*N, N, N', N'*-tetraisopropyl phosphorodiamidite. It can be activated with tetrazole (Zhao and Landry, 1993; Meyer and Hanna, 1996) or tetrazole diisopropyl ammonium salt (Picken, pers. comm.). A method of phosphitylation using 2-cyanoethyl-*N, N, N', N'*-tetraisopropyl phosphorodiamidite and pyridinium trifluoroacetate as the activator, has been developed (Sanghvi *et al.*, 2000). Phosphoramidites are extremely reactive species and are very sensitive to air oxidation; the  $P^{III}$  species are easily oxidised to  $P^V$ .

Three different methods of making the phosphoramidite were used to determine which method was the most efficient and yielded the purest product. Method A (Meyer and Hanna, 1996) used tetrazole as the activator in the presence of 2-cyanoethyl tetraisopropylphosphorodiamidite in a 1:1 ratio, added in two separate portions to prevent over-activation. Method B (Picken, pers. comm.) used 2-cyanoethyl tetraisopropylphosphorodiamidite with 0.3 equivalents of the activator tetrazole diisopropyl ammonium salt. Method C (Sinha *et al.*, 1983) used the phosphoramidite reagent 2-cyanoethyl-*N, N'*-diisopropylchloro phosphine with 2 equivalents of diisopropylethylamine.

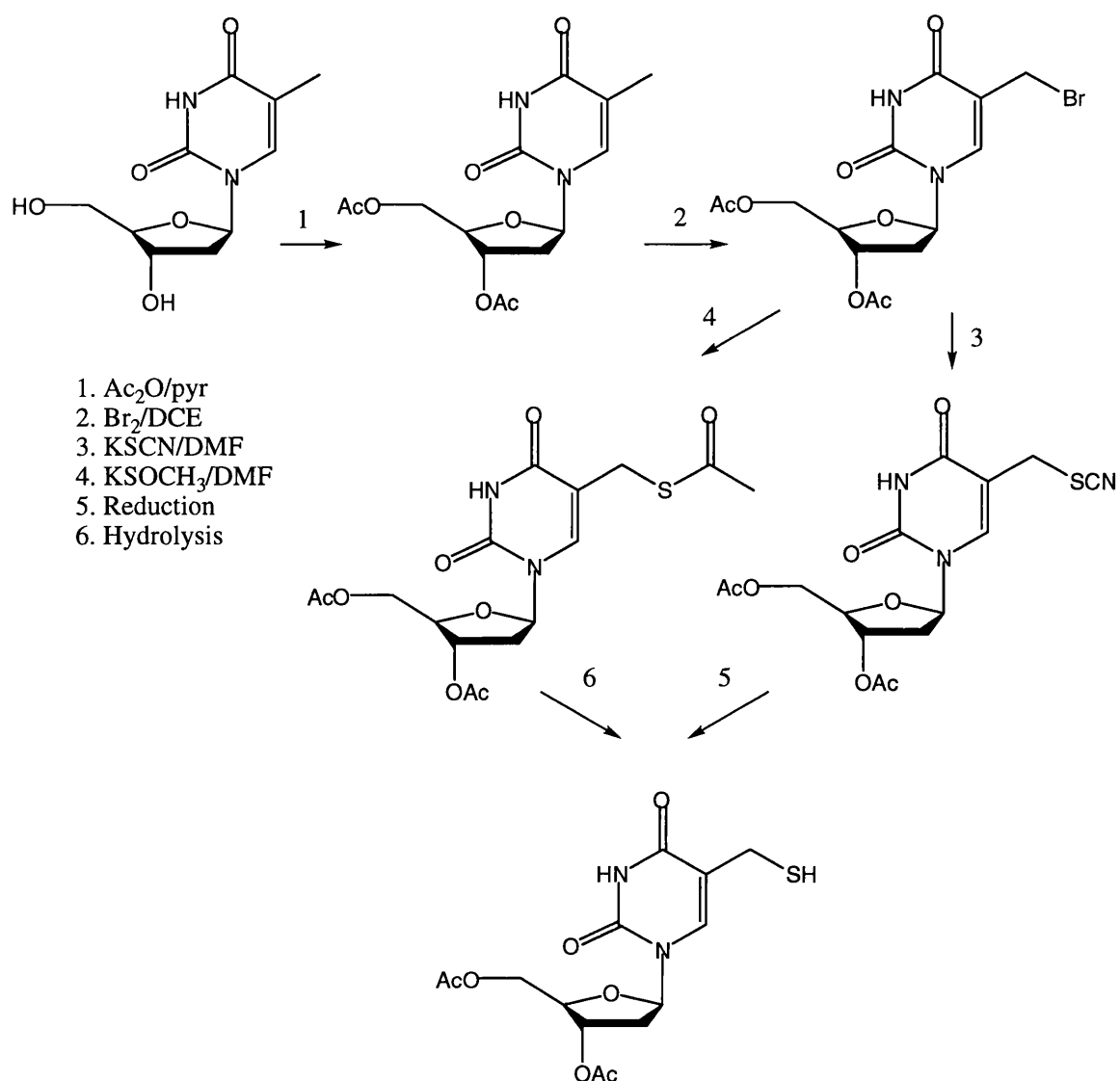
## 2.7 Synthetic routes to thiol targets

The compounds shown in Figure 2.1 are the synthetic targets. Requirements for the DNA synthesis are a trityl-protected primary alcohol (5'-OH) and a cyanoethyl phosphoramidite at the secondary hydroxyl (3'-OH) of the sugar.



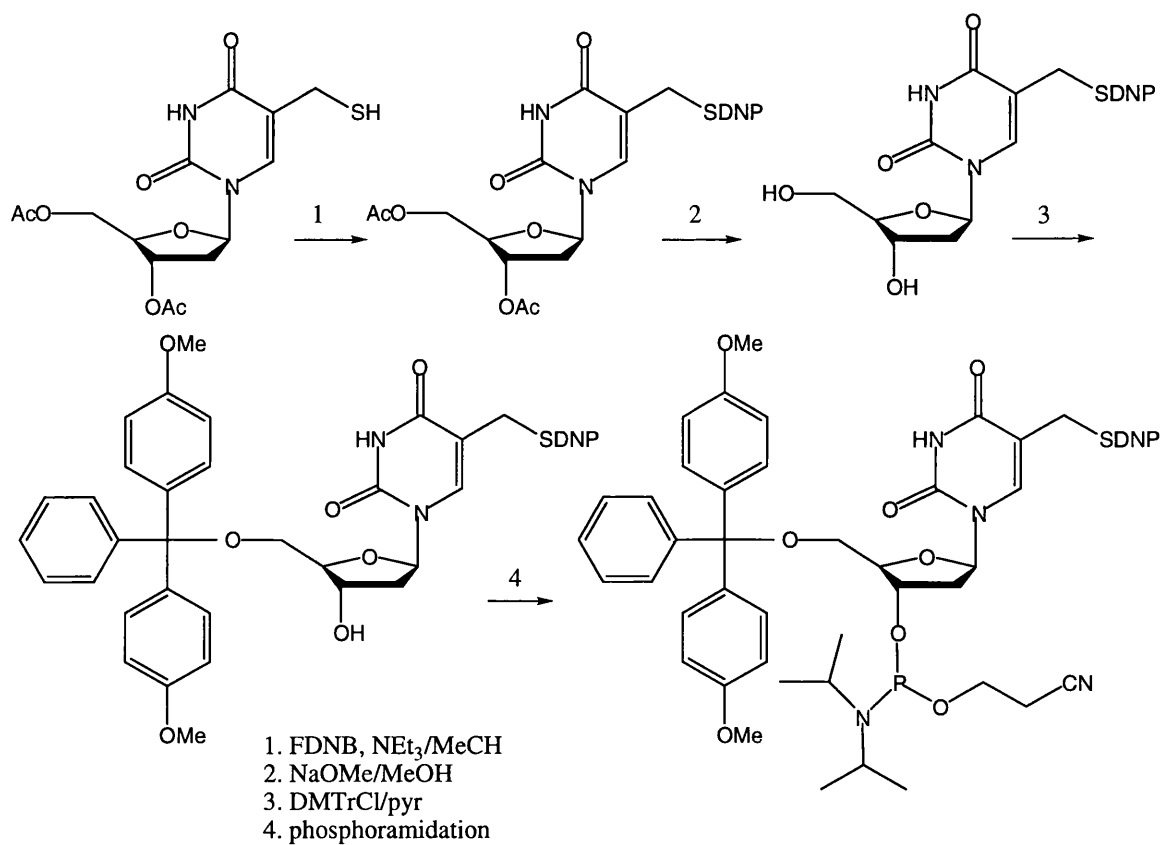
**Figure 2.9 Synthetic route to target A**

Illustration of the route used to synthesise the 5-DNP-protected thiol phosphoramidite



**Figure 2.10 Synthetic route to B**

Illustration of the synthesis of the 5-mercaptomethyl derivative, starting from thymidine.



**Figure 2.11 'Proposed' synthetic route to B**

Diagram showing the protection and the phosphoramidation of the 5-mercaptomethyl derivative.

These protecting groups are easily removed during DNA synthesis, and allow for propagation of the oligonucleotide chain.

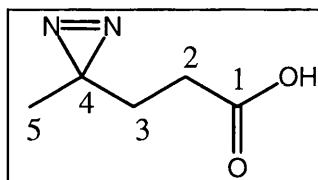
The route to target A (Figure 2.1.2) is shown in Figure 2.9 (Meyer and Hanna, 1996) and the proposed route to target B (Figure 2.1.3) is shown in Figure 2.10 and Figure 2.11.

## 2.8 Experimental

Reagents were purchased from Aldrich Chemical Company (Gillingham, UK), Avocado Synthesis (UK) or Lancaster Synthesis (UK) and were used without further purification. Organic solvents were obtained from Rhône-Poulenc-Rorer and were dried, where necessary, using the procedures described by Leonard *et al.*, 1989. Melting points were recorded in open capillaries using a Gallenkamp apparatus and are uncorrected.  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were recorded for solutions in  $\text{CDCl}_3$  with tetramethylsilane as an internal standard on a Bruker DPX/400 spectrometer operating at 400 and 100 MHz respectively, unless otherwise stated.  $^{13}\text{C}$  NMR spectra were assigned with the aid of Distortionless Enhancement by Polarisation Transfer (DEPT)-edited spectra. All coupling constants are measured in Hz. Thin layer chromatography was performed using Merck aluminium-backed silica plates of 0.25 mm thickness. Chromatograms were visualised using UV illumination at 254 nm, staining with iodine or using a variety of common stains prepared by the methods described in Leonard *et al.*, 1989. Column chromatography was carried out on silica gel (particle size 70-230 mesh). Mass spectra (MS) were recorded on AEI MS12 or MS902 spectrometers using the electron-impact ionisation (EI) mode or, if stated, chemical ionisation (CI) or fast atom bombardment (FAB) modes. Infra-red (IR) spectra were recorded on a Perkin Elmer PU 9800 FT-IR spectrometer. Combustion analysis was carried out on a Carlo-Erba 1106 elemental analyser.

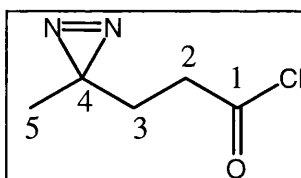


4, 4-Azopentanoic acid **SH P1** (Church and Weiss, 1970)



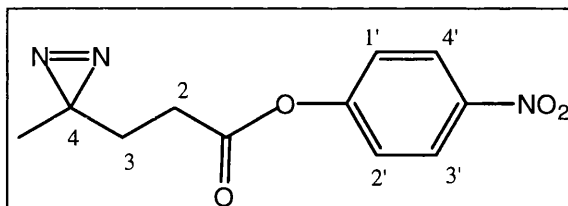
A solution of 4-ketopentanoic acid (3.17 g, 27.2 mmol) in liquid ammonia (40 ml) was stirred for 5 h at reflux. The solution was cooled to  $-78^{\circ}\text{C}$  using an acetone/dry  $\text{CO}_2$  bath, and a suspension of hydroxylamine-*O*-sulphonic acid (3.8 g, 33.6 mmol) in dry methanol (10 ml) was added over a period of 30 min. The mixture was stirred at reflux for 1 h, the ammonia was then evaporated overnight. The resulting slurry was filtered and the solid washed with methanol ( $5 \times 10$  ml). All washings were combined with the original filtrate and the solution was concentrated *in vacuo* to about 10 ml. Excess ammonia was removed using an oil pump ( $< 0.1$  mmHg) until no odour was detected. The crude methanolic solution of 4,4-hydrazipentanoic acid was diluted with dry methanol (20 ml), cooled in an ice bath and treated with triethylamine (3.46 ml, 24.8 mmol) and solid iodine (3.6 g, 14.2 mmol) was added over 1 h. After all the iodine was added, a red colour of excess iodine persisted. The solution was concentrated *in vacuo* to 20 ml and then diluted with brine (50 ml). The solution was extracted with diethyl ether ( $3 \times 50$  ml), and the combined extracts were dried using  $\text{MgSO}_4$ , filtered and concentrated *in vacuo* yielding the title compound as a brown oil (509.7 mg, 14.6%). The residue was rigorously dried using an oil pump ( $< 0.1$  mmHg). New data: (Found C, 46.88; N, 21.78; H, 6.38.  $\text{C}_5\text{H}_8\text{N}_2\text{O}_2$  requires C, 46.89; N, 21.86; H, 6.25%);  $\nu_{\text{max}}$  (liquid film, NaCl plates)/ $\text{cm}^{-1}$  2928m (OH), 1715s (CO), 1585s (N=N);  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) 0.97 (3H, s,  $\text{CH}_3$ -5), 1.63 (2H, t,  $J=7.8$ ,  $\text{CH}_2$ -3), 2.16 (2H, t,  $J=7.8$ ,  $\text{CH}_2$ -2), 7.19 (1H, s, OH);  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) 18.7 ( $\text{CH}_3$ -5), 27.6 ( $\text{CH}_2$ -3), 28.3 ( $\text{CH}_2$ -2), 44.5 (C-4), 177.3 (CO);  $m/z$  ( $\text{Cl}^+$ ) 129.0664 ( $\text{M}+\text{H}^+$ )  $\text{C}_5\text{H}_9\text{N}_2\text{O}_2$  requires 129.0663;  $m/z$  (%) 129.11 (32), 101.1 (100), 83 (85).

4,4-Azopentanoyl chloride **SH P2** (Church and Weiss, 1970)



A solution of 4,4 azopentanoic acid (235 mg, 1.83 mmol) and oxalyl chloride (0.192 ml, 2.20 mmol) was stirred for 16 h under nitrogen. The excess oxalyl chloride was removed *in vacuo* yielding the title compound as an oil (88%) which was difficult to purify. It was difficult to get rid of the excess oxalyl compound as the product is too volatile and viscous to purify by vacuum distillation.

*p*-Nitrophenyl 4,4-azopentanoate **SH P3** (Endo *et al.*, 1982)

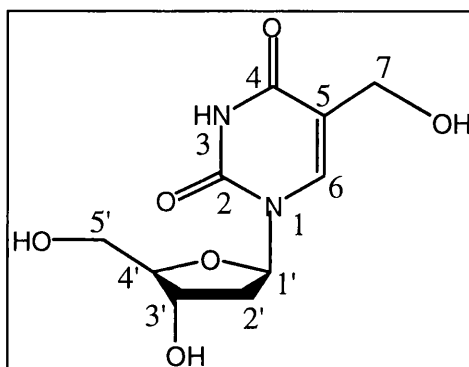


A solution of 4,4-azopentanoic acid (234.8 mg, 1.83 mmol), DCC (0.49 g, 2.38 mmol), DMAP (43.9 mg, 0.36 mmol) and 4-nitrophenol (0.25 g, 1.83 mmol) in dichloromethane (5 ml) was stirred overnight at room temperature. Solvents were removed *in vacuo* yielding a yellow solid that was further purified using silica chromatography eluting with chloroform. The product was collected and concentrated *in vacuo* yielding a clear gum, (40 mg, 8.7%).

New data:  $R_f = 0.47$  (chloroform);  $\nu_{\max}$  (liquid film, NaCl plates)/ $\text{cm}^{-1}$  2925s (CH), 1765s (CO), 1592s (N=N), 1347s ( $\text{NO}_2$ ) and 864s (Ar-H);  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) 1.03 (3H, s, H-5), 1.80 (2H, t,  $J = 7.4$ ,  $\text{CH}_3$ -3), 2.40 (2H, t,  $J = 7.4$ ,  $\text{CH}_2$ -2), 7.22 (2H, m, H1' and H2'), 8.20 (2H, m, H3' and H4');  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) 20.1 ( $\text{CH}_3$ -5), 29.3 ( $\text{CH}_2$ -3), 29.7 ( $\text{CH}_2$ -2), 60.7 (C-4), 122.7 (CH-3' and CH-4'), 125.6 (CH-1' and CH-2'),

138.5 (C-O), 155.6 (C-NO<sub>2</sub>) and 170.3 (CO);  $m/z$  (Cl<sup>+</sup>) 250.0829 (M+H)<sup>+</sup>  
 C<sub>12</sub>H<sub>12</sub>N<sub>3</sub>O<sub>4</sub> requires 250.0823;  $m/z$  (%) 250.1 (100), 222.1 (55), 140 (30).

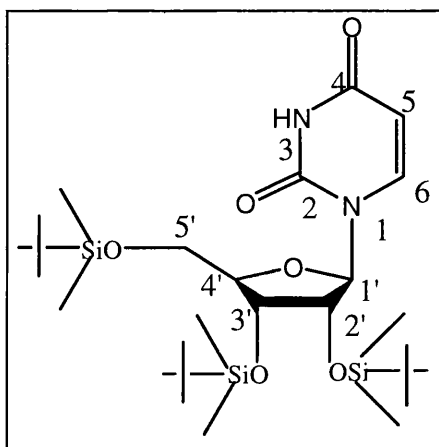
5-Hydroxymethyl-2'-deoxyuridine **SH P4** (Shiau *et al.*, 1980)



Paraformaldehyde (260 mg, 8.9 mmol) was added to a solution of 2'-deoxyuridine (200 mg, 0.87 mmol) in potassium hydroxide (0.5 M; 5 ml). The solution was heated to 60 °C for 6 d. After 1 and 3 d the pH of the solution was adjusted to 11 using potassium hydroxide (0.5M, 2 ml). The reaction mixture was cooled and neutralised using freshly activated Dowex 50WX8(H<sup>+</sup>) resin for 1 h. The solution was filtered and concentrated *in vacuo* yielding the crude product that was purified using column chromatography on silica, eluting with 4:1 chloroform:methanol. The title compound was obtained as white prisms (18 mg, 7.9%); mp = 177-180 °C (Shiau *et al.*, mp = 179-180 °C);  $R_f$  = 0.28 (4:1 chloroform:methanol); (Found C, 46.15; H, 5.32; N, 10.93. C<sub>10</sub>H<sub>14</sub>N<sub>2</sub>O<sub>6</sub> requires C, 46.53; H, 5.42; N, 10.85%);  $\delta_H$  (CDCl<sub>3</sub>) 2.51 (1H, m, H-2'), 2.50 (1H, m, H-2''), 3.55 (1H, dd, J = 11.9 and 3.3, H5'), 3.56 (1H, dd, J = 11.9 and 3.3, H5''), 3.84 (1H, m, H-4'), 4.12 (2H, m, CH<sub>2</sub>-5), 4.30 (1H, dd, J = 3.9 and 1.9, H-3'), 4.95 (1H, m, CH<sub>2</sub>OH), 5.25 (1H, m, OH-5'), 5.40 (1H, d, J = 5.4, OH-3'), 5.80 (1H, dd, J = 5.8, H-1'), 7.45 (1H, s, J = 7.89, H-6), 11.40 (1H, br. s, NH);  $m/z$  (Cl<sup>+</sup>) 259.2325 (M+H)<sup>+</sup> C<sub>10</sub>H<sub>15</sub>N<sub>2</sub>O<sub>6</sub> requires 259.2310;  $m/z$  (%) 259.2 (14), 201.1 (8) and 81.0 (100).

New data:  $\nu_{\max}$ (KBr disc)/ $\text{cm}^{-1}$  3506s (OH), 3045w (CH), 1775s and 1690s (CO) and 1600s (C=C);  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) 56.3 ( $\text{CH}_2$ -2'), 61.4 ( $\text{CH}_2$ -7), 63.2 ( $\text{CH}_2$ -5'), 70.4 (CH-4'), 73.6 (CH-3'), 89.7 (CH-1'), 114.5 (C-5), 137.4 (C-6), 151.0 (CO), 162.9 (CO).

2',3',5'-Bis -O-(*t*-butyldimethylsilyl)-uridine **SH P5** (Ogilvie *et al.*, 1978)

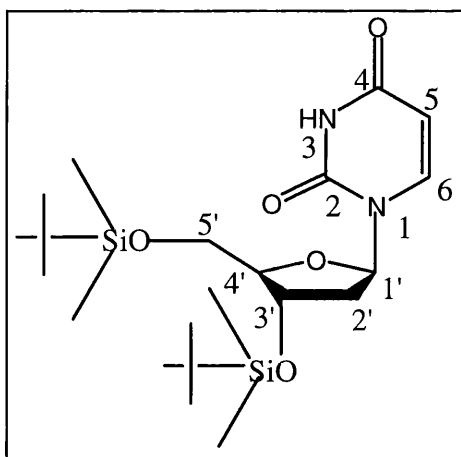


A solution of uridine (0.50 g, 2.04 mmol), *tert*-butyldimethylsilylchloride (2.16 g, 14.3 mmol) and imidazole (1.95 g, 28.5 mmol) in dry dimethylformamide (5 ml) was stirred for 8 h at room temperature. The solvent was removed *in vacuo*, the residue dissolved in a minimum amount of chloroform, and purified using column chromatography on silica, eluting with 2:1 ether:hexane. The relevant fractions were pooled yielding the title compound as a white gum (1.12 g, 93.0%).

New data: mp=76.2-78.4 °C;  $R_f$ = 0.56 (2:1 ether:hexane); (Found C, 54.60; H, 9.27; N, 4.74.  $\text{C}_{27}\text{H}_{54}\text{N}_2\text{O}_6\text{Si}_3$  requires C, 55.28; H, 9.20; N, 4.77%);  $\nu_{\max}$  (KBr disc)/ $\text{cm}^{-1}$  2955 and 2930w (CH), 1695s (CO), 1274s (SiC), 836s (SiO);  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) -0.03, -0.02, 0.00 and 0.01 (18H, 6  $\times$  s, 6  $\times$  SiMe), 0.78 and 0.81 (27H, 3  $\times$  s, 3  $\times$  *t*-butyl), 3.67 (1H, d,  $J$  = 11.5, H-5'), 3.90 (1H, d,  $J$  = 11.5, H-5''), 3.98 (1H, d,  $J$  = 3.0, H-3'), 3.99 (1H, m, H-4'), 4.00 (1H, d,  $J$  = 3.4, H-2'), 5.59 (1H, d,  $J$  = 8.1Hz, H-5), 5.77 (1H, m, H-1'), 7.79 (1H, d,  $J$  = 8.1, H-6), 8.89 (1H, br. s, NH).  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) -5.2 to -3.8 (6  $\times$  SiMe), 18.3, 18.4 and 18.8 (3  $\times$  qC), 26.1, 26.2 and 26.3 (3  $\times$  *t*-butyl  $\text{CH}_3$ ), 62.1 (C-5), 71.1 (C-4), 76.5 (C-2), 84.8 (C-3), 89.3

(C-1), 102.2 (C-5), 140.7 (C-6), 150.5 (CO) and 163.6 (CO);  $m/z$  ( $\text{Cl}^+$ ) 587.3373 ( $\text{M} + \text{H}^+$ )  $\text{C}_{27}\text{H}_{54}\text{N}_2\text{O}_6\text{Si}_3$  requires 587.3368  $m/z$  (%) 587.7 (100), 529.2 (40), 343.2 (30).

3',5'-Bis -*O*-(*t*-butyldimethylsilyl)-2'-deoxyuridine **SH P6** (Ogilvie *et al.*, 1978; Ahmadian and Bergstrom, 1998)

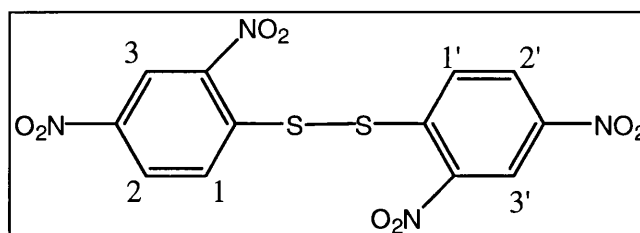


A solution of 2'-deoxyuridine (0.20 g, 0.87 mmol); *tert*-butyldimethylsilyl-chloride (0.92 g, 6.13 mmol) and imidazole (0.83 g, 12.2 mmol) in dry dimethylformamide (3 ml) was stirred for 8 h at room temperature. The solvent was removed *in vacuo* and the residue was dissolved in a minimum amount of chloroform, then purified using silica chromatography with 2:1 ether:hexane as eluant. The relevant fractions were pooled yielding the title compound as a white gum (0.38 g, 95.0%); (Found C, 54.87; H, 8.88; N, 5.80.  $\text{C}_{21}\text{H}_{40}\text{N}_2\text{O}_5\text{Si}_2$  requires C, 55.26; H, 8.76; N, 6.13%);  $\nu_{\text{max}}$  (KBr disc)/ $\text{cm}^{-1}$  2955 and 2929w (CH), 1704s (CO), 1277s (SiC), 837s (SiO);  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) -0.03, -0.02, 0.00 and 0.01 (12H, 4  $\times$  s, 4  $\times$  SiMe), 0.78 and 0.81 (18H, 2  $\times$  s, 2  $\times$  *t*-butyl), 1.95 (1H, m, H-2'), 2.22 (1H, m, H-2''), 3.64 (1H, m, H-4'), 3.81 (3H, m, H-4' and H-5'), 4.31 (1H, m, H-3'), 5.57 (1H, d,  $J = 8.1\text{Hz}$ , H-5), 6.18 (1H, m, H-1'), 7.79 (1H, d,  $J = 8.1$ ), 8.28 (1H, br. s, NH).

New data: mp=110-112 °C;  $R_f = 0.41$  (2:1 ether:hexane); (Found C, 54.87; H, 8.88; N, 5.80.  $\text{C}_{21}\text{H}_{40}\text{N}_2\text{O}_5\text{Si}_2$  requires C, 55.26; H, 8.76; N, 6.13%);  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) -

5.2 to -4.2 ( $4 \times \text{SiMe}$ ), 18.3, 18.7 ( $2 \times \text{qC}$ ), 26.0 and 26.2 ( $2 \times \tau\text{-butyl CH}_3$ ), 42.3 ( $\text{CH}_2\text{-2'}$ ), 62.8 ( $\text{CH}_2\text{-5'}$ ), 71.5 ( $\text{CH-4'}$ ), 85.5 ( $\text{CH-3'}$ ), 88.1 ( $\text{CH-1'}$ ), 102.4 ( $\text{CH-5}$ ), 140.5 ( $\text{CH-6}$ ), 150.3 ( $\text{CO}$ ) and 163.3 ( $\text{CO}$ );  $m/z$  ( $\text{Cl}^+$ ) 457.2555 ( $\text{M} + \text{H}^+$ )  $\text{C}_{21}\text{H}_{41}\text{N}_2\text{O}_5\text{Si}_2$  requires 457.2556  $m/z$  (%) 457.5 (100), 213.3 (22), 145.3 (44).

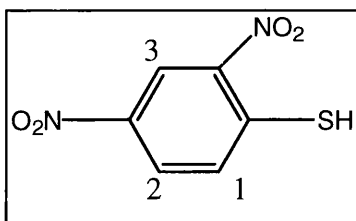
*bis*-(2,4-dinitrophenyl)disulfide **SH P7** (Kharasch *et al.*, 1955)



A solution of 2,4-dinitrobenzenesulfonyl chloride (1 g, 4.26 mmol) in glacial acetic acid (14.7 ml) was warmed until complete dissolution occurred, then allowed to stir for a further 5 min. After the addition of water (9.8 ml), a yellow precipitate ensued. Further water (14.7 ml) was added to the reaction vessel, and the mixture was stirred for 20 min. The yellow solid was collected by vacuum filtration, and dried overnight *in vacuo* using  $\text{P}_2\text{O}_5$ . The crude yellow solid was heated at reflux in freshly dried (distilled from  $\text{CaH}_2$ ) benzene (20 ml). The insoluble material was collected by filtration, washed with diethyl ether and dried over  $\text{P}_2\text{O}_5$  *in vacuo* yielding the title compound as a pale yellow solid (1.43 g, 84.3%); mp = 235-239 °C (dec.) (Kharasch *et al.*, 1955, mp = 240-280 °C).

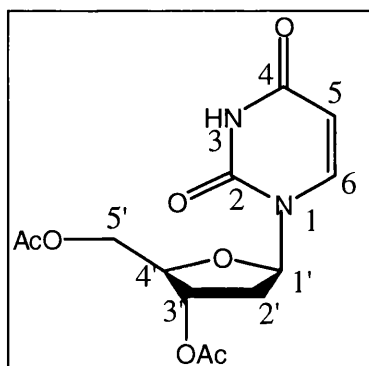
New data:  $R_f$  = 0.56 (2:1 ether:hexane);  $\nu_{\text{max}}$  (KBr)  $\text{cm}^{-1}$  1598 (Ar), 1535 and 1350 ( $\text{NO}_2$ ) stretch and 809 (*p*-disubstituted Ar);  $\delta_{\text{H}}$  (DMSO) 7.20 (2H, d,  $J$  = 8.4, H-1 and H-1'), 8.24 (2H, dD,  $J$  = 1.2, 8.4, H-2 and H-2') and 8.87 (2H, d,  $J$  = 1.2, H-3 and H-3');  $m/z$  ( $\text{EI}^+$ ) 397.9625 ( $\text{M}^+$ )  $\text{C}_{12}\text{H}_6\text{N}_4\text{S}_2$  requires 397.9630;  $m/z$  (%) 398.0 (12), 199.0 (100), 95.0 (52) and 63.0 (87).

2,4-Dinitrobenzenethiol **SH P8** (Bordwell and Andersen, 1953; Kharasch and Parker, 1959)



The disulfide **SH P7** (4.26 g, 10.7 mmol) and glucose.H<sub>2</sub>O (5.94 g, 30 mmol) in ethanol (15 ml) was heated to 60 °C. Once the temperature was reached, NaOH (2.4 g, 60 mmol) in H<sub>2</sub>O (6 ml) was added dropwise. The mixture was kept at 60 °C for 15 min, then the reaction was diluted with H<sub>2</sub>O (to 100 ml) and filtered into HCl and ice (10 ml). The mixture was kept on ice until a orange-brown precipitate appeared. The solid was collected by filtration, dissolved in water and extracted with chloroform. The organic layer was dried with MgSO<sub>4</sub>, and concentrated *in vacuo* yielding the crude product (530 mg, 24.7%). This was purified using column chromatography on silica eluting with 4:1 chloroform:methanol yielding the title compound as a red solid (70 mg, 13.2% from crude); mp = 132-134 °C (Kharasch and Parker, 1959, mp = 128-130 °C). New data; R<sub>f</sub> = 0.27 (4:1 chloroform:methanol); ν<sub>max</sub> (KBr )/cm<sup>-1</sup> 2253 (SH), 1598 (Ar), 1529 and 1342 (NO<sub>2</sub>) stretch and 898 (*p*-disubstituted Ar); δ<sub>H</sub> (DMSO) 7.56 (1H, d, J = 8.7, H-1), 8.20 (1H, dd, J = 8.7, 2.4, H-2) and 9.05 (1H, d, J = 2.4, H-3); *m/z* (EI<sup>+</sup>) 199.9893 (M<sup>+</sup>) C<sub>6</sub>H<sub>3</sub>N<sub>2</sub>S requires 199.9892; *m/z* (%) 199.98 (30), 149.0 (15), 106.0 (15) and 82.9 (100).

3',5'-*O*-Diacetyl-2'-deoxyuridine **SH P9** (Lin and Gao, 1983)

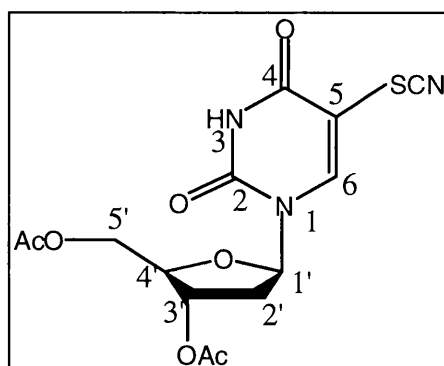


2'-Deoxyuridine (2.2 g, 9.6 mmol) was evaporated three times with anhydrous pyridine (3 × 10 ml) then re-dissolved in pyridine (20 ml) and freshly distilled acetic anhydride (10.7 ml, 96.9 mmol) and imidazole (1.86 g, 13.6 mmol) were added. The reaction was stirred overnight at 40 °C. The pyridine was removed *in vacuo* and the residue was azeotroped with toluene (3 × 15 ml). The resulting oil was recrystallised from ethanol yielding the product as white needles (2.51 g, 82.3%); mp = 126-128 °C (Lin and Gao, 1983, mp = 106-108 °C);  $\delta_{\text{H}}(\text{CDCl}_3)$  2.05 and 2.04 (3H, 2 × s, OCH<sub>3</sub>-5' and OCH<sub>3</sub>-3'), 2.11 (1H, m, H-2'), 2.48 (1H, ddd, J = 14.2, 5.6 and 1.9, H-2''), 4.20 (1H, ddd, J = 6.7, 5.9 and 3.9, H-4'), 4.29 (1H, dd, J = 12.1 and 4.2, H-5') and 4.30 (1H, dd, J = 12.1 and 4.2, H-5''), 5.15 (1H, dd, J = 6.4 and 6.4, H-3'), 5.72 (1H, d, J = 8.2, H-5), 6.22 (1H, dd, J = 5.6 and 5.6, H-1'), 7.43 (1H, d, J = 8.2, H-6) and 9.58 (1H, br. s, NH).

New data:  $R_f$  = 0.7 (4:1 CHCl<sub>3</sub>:MeOH); (Found C, 49.99; H, 5.12; N, 8.93.

C<sub>13</sub>H<sub>16</sub>N<sub>2</sub>O<sub>7</sub> requires C, 50.02; H, 5.13; N, 8.97%);  $\nu_{\text{max}}(\text{KBr disc})/\text{cm}^{-1}$  3183m (NH), 3062w (CH), 1745s and 1689s (CO) and 1624s (C=C);  $\delta_{\text{C}}(\text{CDCl}_3)$  21.2 (OCH<sub>3</sub>-5' and OCH<sub>3</sub>-3'), 38.2 (CH<sub>2</sub>-2'), 64.1 (CH<sub>2</sub>-5'), 74.4 (CH-4'), 82.7 (CH-3'), 85.7 (CH-1'), 103.3 (CH-5), 139.2 (CH-6), 150.7 (CO), 163.5 (CO), 170.6 and 170.7 (OCCH<sub>3</sub>-5' and OCCH<sub>3</sub>-3');  $m/z$  (EI<sup>+</sup>) 312.0956 (M<sup>+</sup>) C<sub>13</sub>H<sub>16</sub>N<sub>2</sub>O<sub>7</sub> requires 312.0965;  $m/z$  (%) 312.1 (10), 201.1 (20) and 81.0 (100).

3', 5'-*O*-Diacetyl-5-(thiocyanato)-2'-deoxyuridine **SH P10** (Meyer and Hanna, 1996)



Anhydrous glacial acetic (65 ml) was placed in a three-necked flask with a bubbler delivering dry Cl<sub>2</sub> gas. The gas was bubbled through until a deep yellow

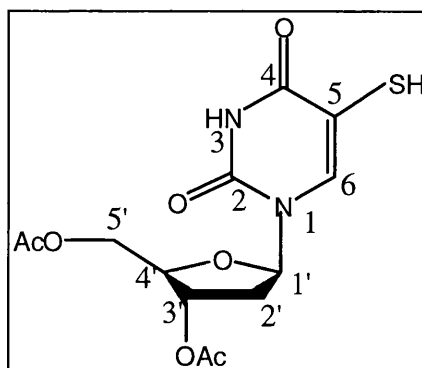


colour persisted after which time a 5 ml aliquot was removed and titrated against a known standard of sodium thiosulfate. To this solution of  $\text{Cl}_2$  (3.68 g, 51.9 mmol) was added potassium thiocyanate (5.57 g, 57.4 mmol) which had been dried overnight at 100 °C. The reaction mixture was stirred at room temperature for 30 min, then 3', 5'-diacetyl-2'-deoxyuridine (1.61 g, 5.18 mmol) **SH P9** was added. The mixture was then stirred for a further 90 min before cyclohexene (10 ml) was added to remove any unreacted  $\text{Cl}_2$ . After stirring for 15 min, the reaction mixture was filtered and the filtrate was concentrated *in vacuo* at  $\geq 40$  °C.

Residual glacial acetic acid was removed azeotropically with toluene and the residue was triturated overnight with 40-60° petroleum ether. The ether was decanted and the residue was dried briefly on a rotary evaporator. The crude mix was further purified using column chromatography on silica eluting with 20:1 ethyl acetate:acetone yielding the title compound as an off-white solid which was recrystallised from ethanol (1.41 g, 73.5%); mp = 151-152 °C (Meyer and Hanna, 1996, mp = 154-156 °C);  $\nu_{\text{max}}$  (KBr disc)/ $\text{cm}^{-1}$  3083m (NH), 2161s (SCN), 1730s and 1690s (CO) and 1619s (C=C);  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) 2.06 and 2.11 (3H, 2  $\times$  s,  $\text{OCH}_3$ -5' and  $\text{OCH}_3$ -3'), 2.16 (1H, m, H-2') and 2.56 (1H, ddd,  $J = 14.2, 5.7$  and  $2.1$ , H-2''), 4.27 (1H, ddd,  $J = 6.7, 5.9$  and  $3.9$ , H-4'), 4.39 (1H, dd,  $J = 12.1$  and  $4.2$ , H-5'), 4.40 (1H, dd,  $J = 12.1$  and  $4.2$ , H-5''), 5.16 (1H, ddd,  $J = 6.4, 4.2$  and  $2.1$ , H-3'), 6.16 (1H, dd,  $J = 5.7$  and  $7.8\text{Hz}$ , H-1'), 8.08 (1H, s, H-6), and 9.38 (1H, br. s, NH).

New data:  $R_f = 0.65$  (20:1 ethyl acetate:acetone); (Found C, 45.35; H, 4.18; N, 11.27.  $\text{C}_{14}\text{H}_{15}\text{N}_3\text{O}_7\text{S}$  requires C, 45.55; H, 4.06; N, 11.37%); UV (MeOH)  $\lambda_{\text{max}} = 268$  nm,  $\epsilon = 17,500$   $\text{l mol}^{-1} \text{cm}^{-1}$ ;  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) 21.2 and 21.3 ( $\text{OCH}_3$ -5' and  $\text{OCH}_3$ -3'), 38.9 ( $\text{CH}_2$ -2'), 63.9 ( $\text{CH}_2$ -5'), 74.2 ( $\text{CH}$ -4'), 83.5 ( $\text{CH}$ -3'), 86.8 ( $\text{CH}$ -1'), 109.3 (C-5), 138.5 (qCN), 144.7 ( $\text{CH}$ -6), 149.5 (CO), 159.6 (CO), 170.6 and 170.7 ( $\text{OCH}_3$ -5' and  $\text{OCH}_3$ -3');  $m/z$  ( $\text{CI}^+$ ) 387.0972 ( $\text{M}+\text{NH}_4$ ) $^+$   $\text{C}_{14}\text{H}_{19}\text{N}_4\text{O}_7\text{S}$  requires 387.0978;  $m/z$  (%) 387.1 (100), 217.1 (45) and 200.1 (11).

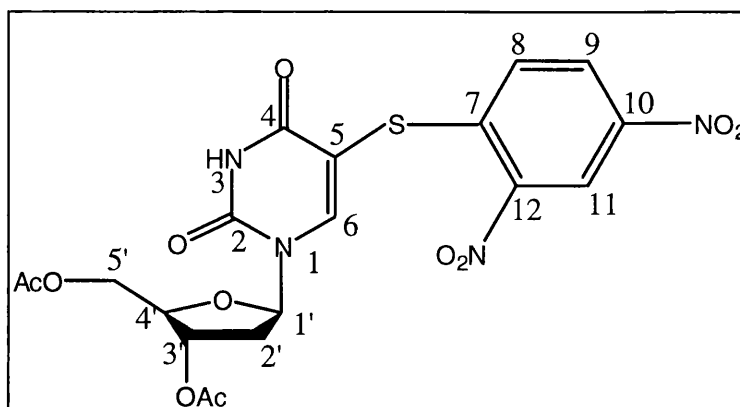
3',5'-*O*-Diacetyl-5-mercapto-2'-deoxyuridine **SH P11** (Meyer and Hanna, 1996)



Ethylenediaminetetraacetic acid (EDTA) disodium salt (169 ml, 0.1M, pH 7.7) and dithiothreitol (2.0 g, 12.9 mmol) were added to a solution of the thiocyanate **SH P10** (1.41 g, 3.8 mmol) in methanol (200 ml). The reaction mixture was stirred at room temperature for 35 min and then filtered. The filtrate was reduced to half volume by rotary evaporation, acidified with 10% (v/v)  $\text{H}_2\text{SO}_4$  and extracted with dichloromethane ( $3 \times 50$  ml). The organic extract was then concentrated *in vacuo* and the residue was triturated with water providing the title compound as a white crumbly solid, (0.98 g, 75.3 %); mp = 153-154 °C (Meyer and Hanna, 1996, mp = 159-163 °C);  $\nu_{\text{max}}$  (KBr disc)/ $\text{cm}^{-1}$  3013m (NH), 2527s (SH), 1744s and 1696s (CO), 1625s (C=C);  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) 2.05 and 2.12 (3H,  $2 \times$  s,  $\text{OCH}_3$ -5' and  $\text{OCH}_3$ -3'), 2.13 (1H, m, H-2') and 2.44 (1H, ddd,  $J = 14.2, 5.7$  and  $2.1$ , H-2''), 3.55 (1H, br.s, SH), 4.21 (1H, dd,  $J = 5.9$  and  $3.0$ , H-4'), 4.30 (1H, ddd,  $J = 12.2, 3.4$  and  $3.0$ , H-5'), 4.31 (1H, ddd,  $J = 12.2, 3.4$  and  $3.0$ , H-5''), 5.16 (1H, ddd,  $J = 6.4, 4.6$  and  $2.2$ , H-3'), 6.25 (1H, dd,  $J = 8.1$  and  $5.7$ , H-1'), 7.60 (1H, s, H-6), and 8.90 (1H, br. s, NH).

New data:  $R_f = 0.26$  (20:1 chloroform:methanol); (Found C, 45.39; H, 4.72; N, 7.95  $\text{C}_{13}\text{H}_{16}\text{N}_2\text{O}_7\text{S}$  requires C, 45.37; H, 4.65; N, 8.14%); UV (MeOH)  $\lambda_{\text{max}} = 260$  nm,  $\epsilon = 16,500 \text{ l mol}^{-1} \text{ cm}^{-1}$  and 320 nm  $\epsilon = 21,000 \text{ l mol}^{-1} \text{ cm}^{-1}$ ;  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) 21.2 and 21.3 ( $\text{OCH}_3$ -5' and  $\text{OCH}_3$ -3'), 38.3 ( $\text{CH}_2$ -2'), 64.2 ( $\text{CH}_2$ -5'), 74.4 (CH-4'), 82.8 (CH-3'), 85.6 (CH-1'), 106.3 (C-5), 138.5 (CH-6), 149.9 (CO), 161.1 (CO), 170.6 and 170.7 ( $\text{OCH}_3$ -5' and  $\text{OCH}_3$ -3');  $m/z$  ( $\text{EI}^+$ ) 344.0681 ( $\text{M}^+$ )  $\text{C}_{13}\text{H}_{16}\text{N}_2\text{O}_7\text{S}$  requires 344.0972;  $m/z$  (%) 344.1 (8), 201.1 (10), 144.0 (12) and 81.0 (100).

3',5'-*O*-Diacetyl-5-[S-(2,4-dinitrophenyl)thio]-2'-deoxyuridine **SH P12** (Meyer and Hanna, 1996)

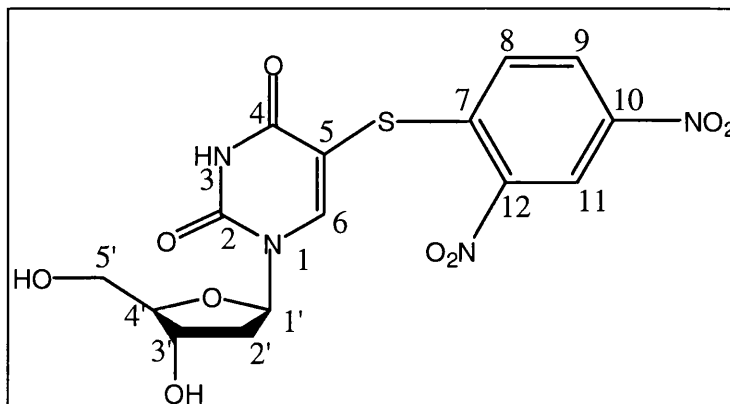


2,4-Dinitrofluorobenzene (134 mg, 0.73 mmol) and triethylamine (486  $\mu$ l) were added to a solution of the mercaptan **SH P11** (200 mg, 0.58 mmol) in anhydrous acetonitrile (24 ml). The reaction mixture was stirred for 90 min at room temperature after which time the solvent was removed *in vacuo*. The crude material was purified using column chromatography on silica eluting with 1:1 hexane:ethyl acetate. The title compound was obtained as shiny yellow flakes (186 mg, 62.8%);  $\nu_{\max}$  (KBr disc)/ $\text{cm}^{-1}$  3447m (NH), 1700s and 1695s (CO), 1616s (C=C), 1560 and 1340 ( $\text{NO}_2$ );  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) 2.00 and 2.04 (3H,  $2 \times$  s,  $\text{OCH}_3$ -5' and  $\text{OCH}_3$ -3'), 2.24 (1H, m, H-2') and 2.53 (1H, ddd,  $J = 14.4$ , 5.8 and 2.4, H-2''), 4.20 (1H, d,  $J = 2.9$ , H-4'), 4.30 (1H, ddd,  $J = 11.0$ , 3.6 and 2.7, H-5'), 4.31 (1H, ddd,  $J = 11.0$ , 3.6 and 2.7, H-5''), 5.15 (1H, dd,  $J = 6.5$  and 3.2, H-3'), 6.18 (1H, dd,  $J = 7.2$  and 6.2, H-1'), 7.27 (1H, d,  $J = 9.8$ , H-8), 8.16 (1H, s, H-6), 8.17 (1H, dd,  $J = 9.9$  and 2.4, H-9) and 8.95 (1H, d,  $J = 2.4$ , H-11).

New data: mp = 85-87  $^{\circ}\text{C}$  (dec.);  $R_f = 0.20$  (1:1 hexane:ethyl acetate; (Found C, 44.40; H, 3.78; N, 10.46  $\text{C}_{19}\text{H}_{18}\text{N}_4\text{O}_{11}\text{S}$  requires C, 44.72; H, 3.53; N, 10.97%); UV (MeOH)  $\lambda_{\max} = 260$  nm,  $\epsilon = 17,300$   $\text{l mol}^{-1} \text{cm}^{-1}$  and 319 nm  $\epsilon = 24,500$   $\text{l mol}^{-1} \text{cm}^{-1}$ ;  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) 19.8 and 20.0 ( $\text{OCH}_3$ -5' and  $\text{OCH}_3$ -3'), 37.2 ( $\text{CH}_2$ -2'), 62.5 ( $\text{CH}_2$ -5'), 72.6 ( $\text{CH}$ -4'), 81.8 ( $\text{CH}$ -3'), 85.1 ( $\text{CH}$ -1'), 102.1 (C-5), 120.6 ( $\text{CH}$ -8), 126.2 ( $\text{CH}$ -9), 127.6 ( $\text{CH}$ -11), 143.4 (C-10), 143.5 (C-12), 144.0 (C-7), 146.2 ( $\text{CH}$ -6), 148.7 (CO), 159.6 (CO), 169.2 and 169.4 ( $\text{OCH}_3$ -5' and  $\text{OCH}_3$ -3');  $m/z$  ( $\text{EI}^+$ )

510.0690 ( $M^+$ )  $C_{19}H_{18}N_4O_{11}S$  requires 510.0695;  $m/z$  (%) 510.1 (8), 201.1 (12) and 81.0 (100).

5-[S-(2,4-Dinitrophenyl)thio]-2'-deoxyuridine **SH P13** (Meyer and Hanna, 1996)

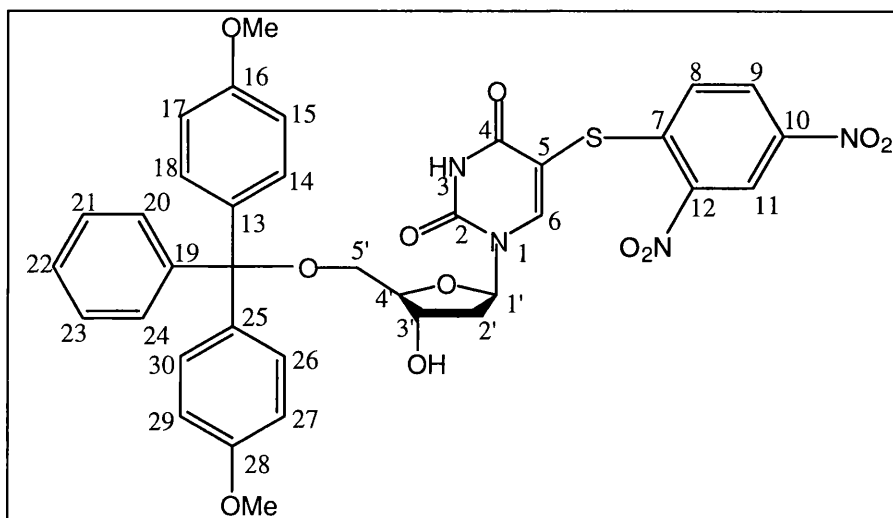


Sodium metal (25 mg, 1.08 mmol) in anhydrous methanol (2 ml) was added to the diacetate **SH P12** (208 mg, 0.408 mmol) in anhydrous methanol (12 ml). The mixture was stirred at room temperature for 2 h, then neutralised with freshly washed Dowex 50WX4 and filtered. The filtrate was concentrated *in vacuo* and the crude product was purified by column chromatography on silica, eluting with ethyl acetate. The title compound was obtained as a yellow solid (150 mg, 86.3%);  $\nu_{\max}$  (KBr disc)/ $\text{cm}^{-1}$  3436m (OH), 1686s (CO), 1596s (Ar), 1521s and 1341s ( $\text{NO}_2$ ) and 1048s (C-OH);  $\delta_{\text{H}}$  (MeOD) 2.26 (1H, m, H-2'), 2.27 (1H, m, H-2''), 3.60 (1H, dd,  $J = 12$  and 3, H5'), 3.69 (1H, dd,  $J = 12$  and 3, H5''), 3.84 (1H, d,  $J = 3.1$ , H-4'), 4.30 (1H, dd,  $J = 3.9$  and 1.9, H-3'), 6.14 (1H, dd,  $J = 6.2$  and 6.2, H-1'), 7.45 (1H, d,  $J = 8.9$ , H-8), 8.20 (1H, dd,  $J = 8.9$  and 2.2, H-9), 8.64 (1H, s, H-6) and 8.87 (1H, d,  $J = 2.4$ , H-11).

New data: mp = 124-126 °C (dec.),  $R_f = 0.19$  (ethyl acetate); UV(MeOH)  $\lambda_{\max} = 267$  nm,  $\epsilon = 16,800 \text{ l mol}^{-1} \text{ cm}^{-1}$  and 325 nm  $\epsilon = 27,500 \text{ l mol}^{-1} \text{ cm}^{-1}$ ;  $\delta_{\text{C}}$  (MeOD) 42.4 ( $\text{CH}_2$ -2'), 62.7 ( $\text{CH}_2$ -5'), 72.1 (CH-4'), 88.0 (CH-3'), 89.7 (CH-1'), 103.3 (C-5), 122.6 (CH-8), 128.6 (CH-9), 130.7 (CH-11), 146.2 (C-10), 146.6 (C-12), 147.1 (C-7), 151.0 (CH-6), 152.2 (CO), 163.9 (CO);  $m/z$  (FAB) 427.0568 ( $M+H$ ) $^+$

$C_{15}H_{15}N_4O_9S$  requires 427.0550;  $m/z$  (%) 426.9 (14), 306.9 (55), 288.9 (30) and 153.9 (100).

5'-O-(Dimethoxytrityl)-5-[S-(2,4-dinitrophenyl)-thio]-2'-deoxyuridine **SH P14**  
(Meyer and Hanna, 1996; D. Picken, pers.comm.)

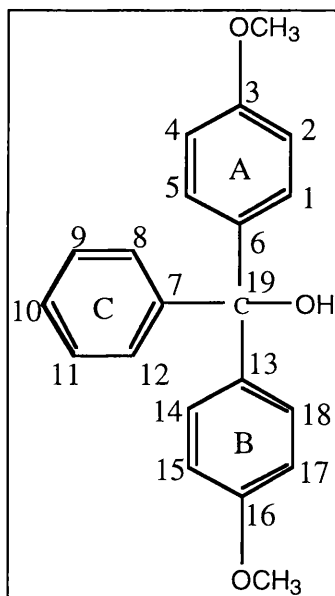


The dihydroxy compound **SH P13** (493 mg, 1.16 mmol) was thoroughly dried using an oil pump then azeotroped with anhydrous pyridine ( $2 \times 5$  ml). To the reaction flask was added 4,4'-dimethoxytrityl chloride (471 mg, 1.39 mmol) and anhydrous pyridine (15 ml). The mixture was stirred overnight at room temperature under argon. The reaction was monitored by tlc and after 48 h stirring, methanol was added to destroy any excess trityl chloride. The solution was concentrated *in vacuo*, yielding a gum which was dissolved in dichloromethane (50 ml) and washed with saturated sodium hydrogen carbonate (30 ml) and brine (30 ml). The organic layer was dried using magnesium sulphate and concentrated *in vacuo*, yielding material which was purified using column chromatography eluting with 9:1 dichloromethane: methanol (and 0.5% triethylamine), yielding the title compound as a yellow gum (710 mg, 84.0%);  $\delta_H$ ( $CDCl_3$ ) 2.32 (1H, m, H-2'), 2.51 (1H, m, H-2''), 3.27 (1H, m, H5'), 3.28 (1H, m, H4''), 3.71 (6H, s,  $2 \times$  OMe), 4.01 (1H, m, H-5'), 4.03 (1H, m, H-5''), 4.56 (1H, m, H-3'), 6.32 (1H, dd,  $J = 7.5$  and  $5.8$ , H-1'), 6.75 (1H, d,  $J = 8.9$ , H-8), 6.99-7.61

(13H,m, trityl group), 7.97 (1H, dd,  $J = 8.9$  and  $2.4$ , H-9), 8.34 (1H, s, H-6) and 8.94 (1H, d,  $J = 2.4$ , H-11);  $m/z$  (FAB) 728.1784 ( $M+H$ )<sup>+</sup>  $C_{36}H_{32}N_4O_{11}S$  requires 728.1790;  $m/z$  (%) 727.9 (10), 303.0 (100), 153.9 (15) and 102.4 (32).

New data: mp =134-138 °C;  $R_f = 0.71$  (9:1 dichloromethane:methanol, 0.5% triethylamine);  $\nu_{max}$  (KBr disc)/cm<sup>-1</sup> 3436m (OH), 2853w (O-Me), 1695s (CO), 1596s (CO), 1508s and 1340s (NO<sub>2</sub>) and 734s (trityl);  $\delta_c$  (CDCl<sub>3</sub>) 45.6 (CH<sub>2</sub>-2'), 63.8 (CH<sub>2</sub>-5'), 73.0 (CH-4'), 86.5 (CH-3'), 87.3 (C-31), 87.5 (CH-1'), 103.3 (CH-5), 113.5 (CH-15, CH-17, CH-27 and 29), 121.8 (CH-8), 127.2 (CH-22), 128.3 (CH-20, CH-21, CH-23 and CH-24), 128.7 (CH-9), 129.5 (CH-11), 130.0 (CH-14, CH-18, CH-26 and CH-30), 138.5 (C-12), 139.8 (C-10), 144.2 (C-13 and C-25), 144.9 (CH-19), 145.9 (C-7), 148.1 (CH-6), 151.5 (C-19), 158.8 (C-16), 158.9 (C-28)and 162.1 (CO).

4,4'-Dimethoxytrityl Alcohol **SH P15** (Bleasdale *et al.*, 1989)

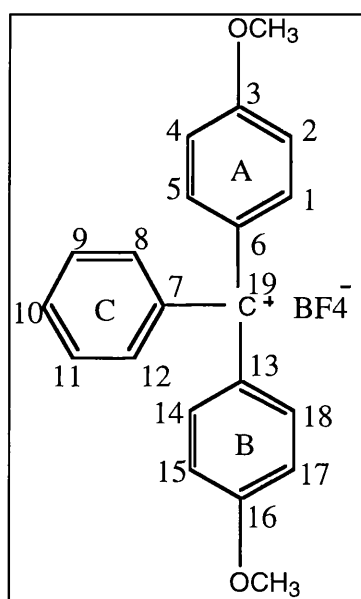


4,4'-Dimethoxytrityl chloride (3.38 g, 10 mmol) in tetrahydrofuran-0.5M aqueous sodium hydroxide (40 ml, 1:1, v/v) was stirred at room temperature for 30 min. The mixture was then extracted with dichloromethane (80 ml), dried using sodium sulfate, filtered and concentrated *in vacuo*. The crude material was purified by column chromatography on silica, eluting with 9:1

dichloromethane:methanol, yielding the title compound as an orange oil which was precipitated from ether-pet.ether (2.36 g, 73.7%); mp = 75-76 °C (Bleasdale *et al.*, 1990, mp = 75-77 °C);  $\delta_{\text{H}}$  (DMSO) 3.71 (6H, s,  $2 \times \text{OCH}_3$ ), 6.22 (1H, s, OH), 6.83 and 7.08 (8H,  $2 \times \text{AA'BB'}$  Ar ring A and B protons) and 7.25 (5H, m, ring C protons).

New data:  $R_f = 0.32$  (dichloromethane); (Found C, 77.78 and H, 6.34  $\text{C}_{21}\text{H}_{20}\text{O}_3$  requires C, 77.76 and H, 6.24%);  $\nu_{\text{max}}$  (KBr disc)/ $\text{cm}^{-1}$  3506s (OH), 2932m (CH) and 1508 and 830 (ArH);  $\delta_{\text{C}}$  (DMSO) 55.3 ( $\text{CH}_3$ -3 and  $\text{CH}_3$ -16), 80.2 (C-19), 113.1 (C-2, 4, 15 and 17), 126.8 (C-10), 127.7 (C-9 and 11), 127.9 (C-8 and 12), 129.2 (C-1, 5, 14 and 18), 140.5 (C-6 and 13), 148.6 (C-7), 158.1 (C-3 and 16);  $m/z$  ( $\text{Cl}^+$ ) 320.1416 ( $\text{M} + \text{H}^+$ )  $\text{C}_{21}\text{H}_{21}\text{O}_3$  requires 320.1413;  $m/z$  (%) 320.1 (45), 243.1 (100), 213.1 (30), 135.0 (63).

4,4'-Dimethoxytrityl Tetrafluoroborate **SH P16** (Bleasdale *et al.*, 1989)



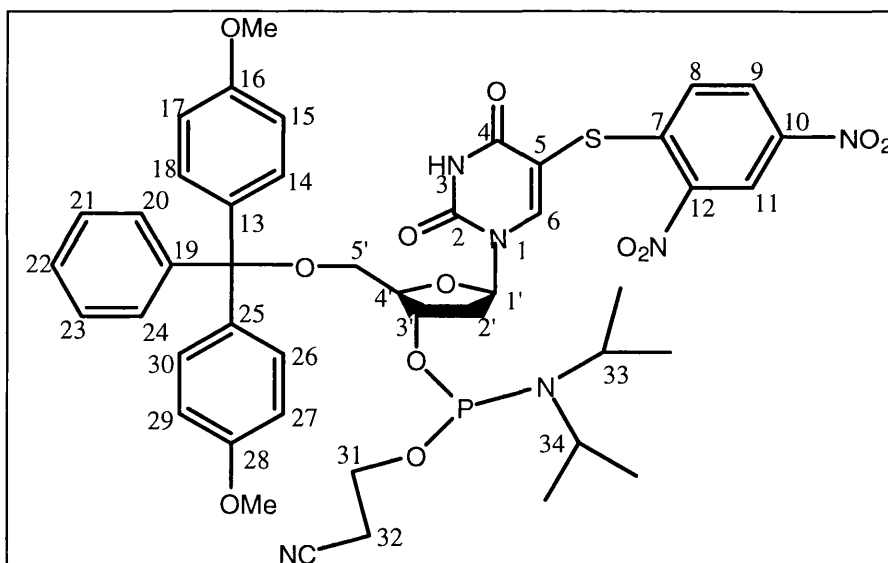
**SH P15** (0.99 g, 3.08 mmol) was dissolved in warm acetic anhydride (6.04 ml, 64 mmol). The solution was allowed to cool, and 40% aqueous tetrafluoroboric acid (1.08 ml, 14.6 mmol) was added at such a rate that the temperature of the reaction mixture did not rise above 25 °C (over 1 h), at which time a dark red solution had formed. Addition of dry ether (100 ml) caused the product to precipitate as deep

orange crystals. These were dried *in vacuo* on a rotary evaporator (20 mm Hg) for 1 h, then using an oil pump (< 0.1 mm Hg) for 2 h to give the title compound as crimson needles, (1.15 g, 96.2%); mp = 165-168 °C (Bleasdale *et al.*, 1990, mp = 193-196 °C);  $\delta_{\text{H}}$  (DMSO) 4.08 (6H, s, 2  $\times$  OCH<sub>3</sub>), 7.20-7.66 (13H, m, rings A, B and C protons);  $m/z$  (EI<sup>+</sup>) 303.1385 (M<sup>+</sup> -BF<sub>4</sub>) C<sub>21</sub>H<sub>19</sub>BF<sub>4</sub>O<sub>2</sub> requires 303.1385;  $m/z$  (%) 303.1 (100), 257.1 (26), 129.1 (28).

New data;  $R_f$  = 0.29 (dichloromethane);  $\nu_{\text{max}}$  (KBr disc)/cm<sup>-1</sup> 3402s (OH), 2925m (CH) and 1508s and 829s (ArH);  $\delta_{\text{C}}$  (DMSO) 57.9 (CH<sub>3</sub>-3 and CH<sub>3</sub>-16), 80.2 (C-19), 117.8 (C-2, 4, 15 and 17), 121.3 (C-10), 122.4 (C-9 and 11), 129.8 (C8 and 12), 130.9 (C-1, 5, 14 and 18), 139.3 (C-6 and 13), 144.8 (C-7), 172.8 (C-3 and 16).

#### Phosphoramidite Method A (Meyer and Hanna, 1996)

5'-*O*-(Dimethoxytrityl)-5-[*S*-(2,4-dinitrophenyl)-thio]-2'-deoxyuridine-3'-*O*-(2-*N,N'*-cyanoethyl diisopropylphosphoramidite **SH P17**



The trityl-protected nucleotide **SH P13** (100 mg, 0.14 mmol) was rigorously dried using an oil pump and kept under dry nitrogen. To the reaction flask was added 2-cyanoethyl tetraisopropylphosphorodiamidite (43  $\mu$ l, 0.13 mmol) and tetrazole (10 mg, 0.13 mmol) in dry acetonitrile (2 ml). The mixture was stirred at



room temperature for 30 min before adding further 2-cyanoethyl-*N, N, N', N'*-tetraisopropylphosphorodiamidite (18  $\mu$ l, 0.05 mmol) and tetrazole (5 mg, 0.06 mmol) in dry acetonitrile (1 ml), and the mixture was stirred for a further 30 min. The reaction was then poured into dichloromethane (20 ml) and 5% (w/v) sodium bicarbonate (15 ml) was added. The organic layer was removed and washed with water, dried with sodium sulfate, filtered and concentrated *in vacuo* yielding a yellow semi-solid. This was further purified using column chromatography on silica eluting with 9:1 dichloromethane:methanol (containing 0.5% triethylamine) yielding the title compound as a yellow gum (41.0 mg, 32.3%).

New data:  $R_f$  = 0.75 (9:1 dichloromethane:methanol);  $\delta_H$  ( $CDCl_3$ ) 1.20 (12H, m, 4  $\times$  iPr), 2.32 (1H, m, H-2'), 2.56 (1H, m, H-2''), 2.68 (4H, dt,  $J$  = 5.9 and  $J$  = 2.3, H-31 and H-32), 3.29 (1H, m, H-5'), 3.30 (1H, m, H-5''), 3.62 (6H, s, 2  $\times$   $OCH_3$ ), 4.01 (1H, m, H-4'), 4.09 (2H, sept, H-33 and H-34), 4.62 (1H, m, H-3'), 6.33 (1H, dd,  $J$  = 7.5 and 5.8, H-1'), 6.75 (1H, d,  $J$  = 8.9, H-8), 7.00-7.24 (13H, m, trityl group), 8.00 (1H, dd,  $J$  = 8.9 and 2.4, H-9), 8.39 (1H, s, H-6) and 8.92 (1H, d,  $J$  = 2.4, H-11).

#### **Phosphoramidite Method B** (D. Picken, pers.commun)

5'-*O*-(Dimethoxytrityl)-5-[*S*-(2,4-dinitrophenyl)-thio]-2'-deoxyuridine-3'-*O*-(2-*N,N'*-cyanoethyl diisopropylphosphoramidite **SH P17**

The trityl-protected nucleoside **SH P13** (554 mg, 0.73 mmol) was dried *in vacuo* with dry acetonitrile (3  $\times$  10 ml). It was then dissolved in dichloromethane (10 ml) and 2-cyanoethyl-*N, N, N', N'*-tetraisopropylphosphorodiamidite (288  $\mu$ l, 0.91 mmol) and tetrazole diisopropylammonium salt (64 mg, 0.36 mmol) was added. The solution was stirred under nitrogen overnight at room temperature. Once the tlc had shown that the starting material had been converted into two slower moving species (corresponding to the diastereoisomers about phosphorus), the reaction was worked up by adding saturated sodium hydrogen carbonate (10 ml). The organic layer was separated and washed with brine (2  $\times$  10 ml). The organic layer was then dried with sodium sulfate and evaporated *in vacuo* to an oil. To

precipitate the product, hexane (100 ml) was added to the oil. The precipitate was left to settle and the solvent was decanted. The precipitate was re-dissolved in ethyl acetate (5 ml) and precipitated by adding hexane (50 ml).

The precipitate was re-dissolved in ethyl acetate (1-2 ml) and purified using column chromatography on silica, eluting with ethyl acetate, yielding the phosphoramidite as a yellow oil. The oil was dissolved in anhydrous acetonitrile, then filtered through a 0.22  $\mu\text{m}$  filter to ensure any small particles were removed. Once filtered, the solution was evaporated and dried by acetonitrile co-evaporation ( $\times 2$ ). The oil obtained was placed under high vacuum, yielding the product as a yellow foam (372 mg, 54.4%).

New data:  $R_f = 0.75$  (9:1 dichloromethane :methanol);  $\delta_H$  ( $\text{CDCl}_3$ ) 1.10 (12H, m,  $4 \times \text{iPr}$ ), 2.42 (1H, t,  $J = 6.2$ ,  $\text{CH}_2\text{CN}$ ), 2.48 (1H, dd,  $J = 6.8$  and  $13.6$ , H-2'), 2.64 (1H, dd,  $J = 5.9$  and  $13.6$  Hz, H-2''), 3.35 (2H, m, H-5' and H-5''), 3.52 (2H, m,  $\text{CHiPr}$ ), 3.62 -3.72 (8H, m,  $2 \times \text{OCH}_3$  and  $\text{CH}_2\text{OP}$ ), 4.19 (1H, m, H-4'), 4.73 (1H, m, H-3'), 6.26 (1H, dd,  $J = 6.5$  and  $6.5$ , H-1'), 6.60-7.20 (13H, m, trityl group), 7.33 (1H, s, H-6), 8.03 (1H, ddd,  $J = 11.4$  and  $2.4$ , H-9), 8.42 (1H, d,  $J = 10.8$ ) and 8.93 (1H, d,  $J = 2.4$ , H-11);  $m/z$  (FAB) 929.2937 ( $\text{M}+\text{H}^+$ )  $\text{C}_{45}\text{H}_{50}\text{N}_6\text{O}_{12}$  PS requires 929.2948;  $m/z$  (%) 929.1 (8), 711.0 (6), 619.1 (12), 406.9 (42) and 303.0 (100);  $\delta_P$  (81 MHz,  $\text{CDCl}_3$ ) 148.7, 148.9.

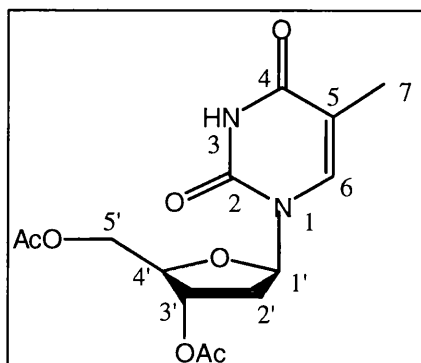
### Phosphoramidite Method C SH P17 (Sinha *et al*, 1983)

The trityl-protected nucleoside **SH P13** (50 mg, 0.07 mmol) was dried *in vacuo* with dry tetrahydrofuran (5 ml). It was then dissolved in tetrahydrofuran (5 ml) and 2-cyanoethyl-*N, N'*-diisopropylchlorophosphorodiamidite (23  $\mu\text{l}$ , 0.10 mmol) and diisopropylethylamine (47  $\mu\text{l}$ , 0.27 mmol) were added. The solution was stirred at room temperature for 4 h, then poured into ethyl acetate (10 ml, pre-treated with  $1 \times 10$  ml and  $3 \times 10$  ml brine washes). The product was precipitated by adding hexane (100 ml). The supernatant was decanted off and the product redissolved in ethyl acetate (10 ml). After adding more hexane (100 ml), the precipitate was collected and dried *in vacuo*. The residue was further purified by

column chromatography on silica eluting with 2:1 ethyl acetate: hexane yielding the product as yellow flakes (18 mg, 28.7%).

New data:  $R_f = 0.70$  2:1 ethyl acetate:hexane;  $\delta_H$  ( $CDCl_3$ ) 1.11 (12H, m,  $4 \times iPr$ ), 2.46 (1H, t,  $J = 6.2$  Hz,  $CH_2CN$ ), 2.51 (1H, dd,  $J = 6.8$  and  $13.6$  Hz, H-2'), 2.65 (1H, dd,  $J = 5.9$  and  $13.6$  Hz, H-2''), 3.49 (2H, m, H5' and H5''), 3.57 (2H, m,  $CHiPr$ ), 3.57 -3.65 (8H, m,  $2 \times OCH_3$  and  $CH_2OP$ ), 4.19 (1H, m, H-4'), 4.74 (1H, m, H-3'), 6.25 (1H, dd,  $J = 6.5$  and  $6.5$ , H-1'), 6.60-7.17 (13H, m, trityl group), 7.32 (1H, s, H-6), 8.03 (1H, ddd,  $J = 11.4$  and  $2.4$ , H-9), 8.43 (1H, d,  $J = 10.8$ ) and 8.93 (1H, d,  $J = 2.4$ , H-11).

3',5'-*O*-Diacetyl-thymidine **SH P18** (Michelson and Todd, 1955)



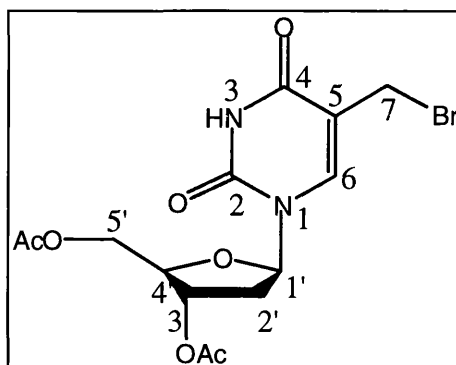
Thymidine (3.2 g, 13.2 mmol) was evaporated three times with dry pyridine ( $3 \times 10$  ml), then dissolved in dry pyridine (20 ml) and freshly distilled acetic anhydride (24 ml, 21.7 mmol) and imidazole (5.2 g, 38.1 mmol) were added. The reaction was stirred overnight at room temperature. The pyridine was removed *in vacuo* and the residue was azeotroped with toluene ( $3 \times 15$  ml). The resulting oil was recrystallised from isopropanol yielding the pure product as white needles (3.5 g, 80.8%); mp  $126-128^\circ C$  (Michelson and Todd, 1955, mp =  $125^\circ C$ ).

New data:  $R_f = 0.7$  (4:1  $CHCl_3$ : MeOH); (Found C, 51.40; H, 5.56; N, 8.43.

$C_{14}H_{18}N_2O_7$  requires C, 51.55; H, 5.52; N, 8.58%);  $\nu_{max}$  (KBr disc)/ $cm^{-1}$  3414m (NH), 3042w (CH), 1750s and 1735s (CO) and 1671s (C=C);  $\delta_H$  ( $CDCl_3$ ) 1.87 (3H, s, H-7), 2.05 and 2.06 (3H,  $2 \times$  s,  $OCH_3$ -5' and  $OCH_3$ -3'), 2.11 (1H, s, H-2'),

2.41 (1H, ddd, J = 14.2, 5.6 and 1.9, H-2''), 4.17 (1H, ddd, J = 6.7, 5.9 and 3.9, H-4'), 4.29 (1H, dd, J = 12.1 and 4.2, H-5'), 4.31 (1H, dd, J = 12.1 and 4.2, H-5''), 5.15 (1H, ddd, J = 6.5, 4.3 and 2.1, H-3'), 6.26 (1H, dd, J = 5.5 and 5.5, H-1'), 7.21 (1H, s, H-6) and 9.19 (1H, br. s, NH);  $\delta_c$  (CDCl<sub>3</sub>) 11.1 (CH<sub>3</sub>-7), 19.8 (OCH<sub>3</sub>-5' and OCH<sub>3</sub>-3'), 36.5 (CH<sub>2</sub>-2'), 62.8 (CH<sub>2</sub>-5'), 73.1 (CH-4'), 81.1 (CH-3'), 83.8 (CH-1'), 110.5 (C-5), 133.4 (CH-5), 149.4 (CO), 162.6 (CO), 170.5 and 169.2 (OCH<sub>3</sub>-5' and OCH<sub>3</sub>-3');  $m/z$  (Cl<sup>+</sup>) 327.1193 (M<sup>+</sup>) C<sub>14</sub> H<sub>18</sub> N<sub>2</sub> O<sub>7</sub> requires 327.1192;  $m/z$  (%) 327.1 (100), 201.1 (27), 127.1 (30), 81.0 (36).

3',5'-O-Diacetyl-5-bromomethyl-2'-deoxyuridine **SH P19** (Sergiev *et al.*, 1997)

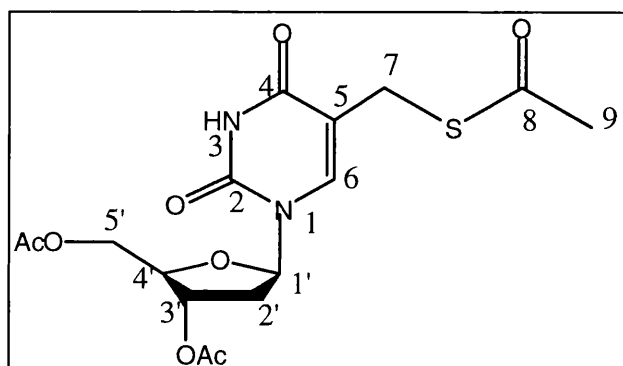


**SH P18** (1.00 g, 3.06 mmol) was dissolved in freshly dried dichloroethane (5 ml, distilled from CaH<sub>2</sub>) in a three necked flask, equipped with a reflux condenser. The mixture was heated to reflux using two powerful tungsten lamps, and Br<sub>2</sub> (0.25 ml, 77.8 mmol) in dichloroethane (2 ml) was added dropwise over 3 h. The mixture was left to cool to room temperature, and concentrated *in vacuo* yielding an orange oil which was azeotroped with toluene (15 ml). The extract was dissolved in the minimum volume of CHCl<sub>3</sub> and purified using silica column chromatography, eluting with 4:1 hexane: ethyl acetate. The relevant fractions were collected and concentrated *in vacuo* yielding a yellow solid, recrystallised by dissolving in the minimum dichloroethane and dropping into hexane, yielding white flakes (550 mg, 44.4%).

New data: mp = 94-98 °C; R<sub>f</sub> = 0.19 (4:1 ethyl acetate:hexane); (Found C, 41.23; H, 4.62; N, 6.51. C<sub>14</sub>H<sub>17</sub>N<sub>2</sub>O<sub>7</sub>Br requires C, 41.51; H, 4.20; N, 6.91%);  $\nu_{\max}$  (KBr

disc)/cm<sup>-1</sup> 3388m (NH), 3080w (CH), 1745s and 1740s (CO), 1691s (C=C) and 742.2 (C-Br);  $\delta_{\text{H}}$  (CDCl<sub>3</sub>) 2.05 and 2.09 (3H, 2 × s, OCH<sub>3</sub>-5' and OCH<sub>3</sub>-3'), 2.08 (1H, m, H-2') and 2.52 (1H, ddd, J = 14.3, 5.5 and 1.7, H-2''), 4.18 (1H, m, H-4'), 4.23 (2H, d, CH<sub>2</sub>-7), 4.32 (1H, m, H-5'), 4.32 (1H, m, H-5''), 5.16 (1H, dd, J = 6.4 and 6.4, H-3'), 6.19 (1H, dd, J = 5.5 and 5.6, H-1'), 7.61 (1H, s, H-6) and 9.38 (1H, br. s, NH);  $\delta_{\text{C}}$  (CDCl<sub>3</sub>) 21.2 (OCH<sub>3</sub>-5' and OCH<sub>3</sub>-3'), 25.2 (CH<sub>2</sub>-7), 38.5 (CH<sub>2</sub>-2'), 64.1 (CH<sub>2</sub>-5'), 74.5 (CH-4'), 83.1 (CH-3'), 86.1 (CH-1'), 111.2 (C-5), 138.8 (CH-5), 150.1 (CO), 161.7 (CO) and 170.7 (OCH<sub>3</sub>-5' and OCH<sub>3</sub>-3');  $m/z$  (CI<sup>+</sup>) 405.0299(M<sup>+</sup>) C<sub>14</sub>H<sub>17</sub>N<sub>2</sub>O<sub>7</sub> Br requires 405.0301;  $m/z$  (%) 407.0 (50), 327.1 (30), 201.1 (30), 127.1 (15).

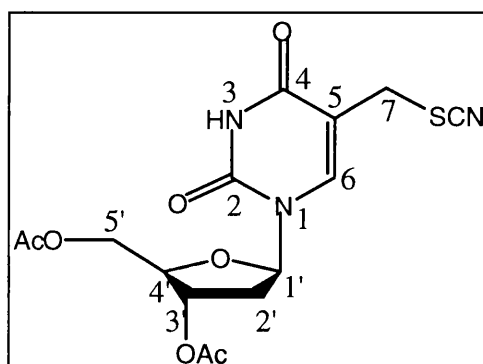
3',5'-O-Diacetyl-5-thioacetylmethyl-2'-deoxyuridine **SH P20** (Langen and Barwolff, 1975)



A solution of **SH P19** (2.0 g, 4.93 mmol) and potassium thioacetate (0.85 g, 7.50 mmol) in dry dimethylformamide (15 ml) was heated to 70 °C under nitrogen and stirred overnight. The reaction mixture was diluted to with H<sub>2</sub>O (50 ml) and extracted with chloroform (2 × 50 ml). The dimethylformamide was removed by washing this organic layer with H<sub>2</sub>O (500 ml). The solution was then dried using anhydrous MgSO<sub>4</sub> and concentrated *in vacuo*, yielding a crude product which was subsequently purified using column chromatography on silica eluting with ethyl acetate. The relevant fractions were pooled and concentrated *in vacuo* yielding product which was recrystallised from isopropanol, yielding the title compound as deep orange flakes (882 mg, 53.7%).

New data: mp = 140-141 °C;  $R_f$  = 0.68 (ethyl acetate); (Found C, 48.08; H, 4.92; N, 6.92.  $C_{16}H_{20}N_2O_8S$  requires C, 48.02; H, 4.99; N, 6.99%);  $\nu_{\max}$  (KBr disc)/ $\text{cm}^{-1}$  3435m (NH), 3077w (CH), 1744s (CO) and 1690s (C=C);  $\delta_H$  ( $\text{CDCl}_3$ ) 2.05, 2.14 and 2.24 (3H,  $3 \times s$ ,  $\text{OCH}_3$ -5',  $\text{OCH}_3$ -3' and S- $\text{OCH}_3$ ; 2.17 (1H, m, H-2''), 2.38 (1H, ddd,  $J$  = 14.2, 5.5 and 1.4, H-2''), 3.70 (2H, dd,  $J$  = 14 and 14, 2H-7), 4.27 (3H, m, H-5', H-5'' and H-4'), 5.18 (1H, dd,  $J$  = 6.4 and 1.8, H-3'), 6.25 (1H, dd,  $J$  = 5.5 and 5.5) and 7.65 (1H, s, H-6);  $\delta_C$  ( $\text{CDCl}_3$ ), 19.9 ( $\text{OCH}_3$ -5' and  $\text{OCH}_3$ -3'), 24.8 ( $\text{CH}_2$ -7), 29.5 ( $\text{OCH}_3$ -9) 36.5 ( $\text{CH}_2$ -2'), 63.0 ( $\text{CH}_2$ -5'), 73.3 ( $\text{CH}$ -4'), 81.3 ( $\text{CH}$ -3'), 83.9 ( $\text{CH}$ -1'), 110.7 (C-5), 136.6 ( $\text{CH}$ -6), 149.0 (CO), 161.3 (CO) and 169.4 ( $\text{OCH}_3$ -5' and  $\text{OCH}_3$ -3');  $m/z$  ( $\text{CI}^+$ ) 400.0937 ( $M^+$ )  $C_{16}H_{20}N_2O_8S$  requires 400.0946;  $m/z$  (%) 400.1 (5), 357.1(12), 201.1 (10), 157.0 (10), 81.0 (100).

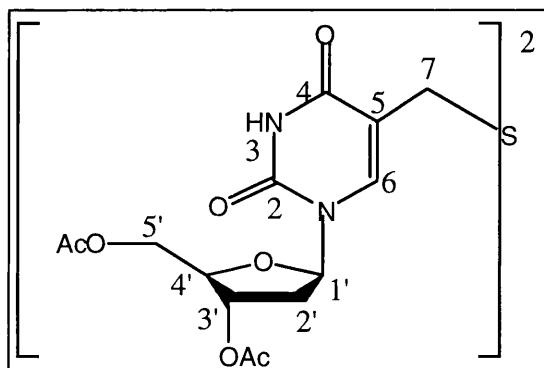
3',5'-*O*-Diacetyl-5-thiocyanatomethyl-2'-deoxyuridine **SH P21** (Langen and Barwolff, 1975)



To 3',5'-*O*-diacetyl-5-bromomethyl-2'-deoxyuridine **SH P19** (993 mg, 2.45 mmol) in anhydrous dimethylformamide (10 ml) was added potassium thiocyanate (490 mg, 5.05 mmol). The reaction mixture was heated to 70 °C and stirred under  $N_2$  overnight. The crude reaction mixture was diluted with  $H_2O$  (50 ml) and extracted with chloroform ( $2 \times 50$  ml). This extract was then washed with  $H_2O$  (1 litre) to remove excess dimethylformamide. The organic layer was dried with  $MgSO_4$  and concentrated *in vacuo* yielding the crude product which was further purified using column chromatography on silica, eluting with ethyl acetate. The title compound was isolated as yellow foam (550 mg, 58.6%).

New data: mp = 53-55 °C,  $R_f$  = 0.52 (ethyl acetate); (Found C, 46.60; H, 4.90; N, 10.46  $C_{15}H_{17}N_3O_7S$  requires C, 47.01; H, 4.50; N, 10.86%);  $\nu_{max}$  (KBr disc)/ $cm^{-1}$  3340m (NH), 3075w (CH), 1740s and 1735s (CO), 1691s (C=C);  $\delta_H$ ( $CDCl_3$ ) 2.06 and 2.07 (3H,  $2 \times$  s,  $OCH_3$ -5' and  $OCH_3$ -3'), 2.14 (1H, m, H-2'), 2.50 (1H, ddd,  $J$  = 14.2, 5.5 and 1.6, H-2''), 4.25 (2H, m, 2H-5), 4.40 (3H, m, H4' and 2 H-5'), 5.18 (1H, dd,  $J$  = 6.3 and 6.4, H-3'), 6.21 (1H, dd,  $J$  = 8.4 and 5.7, H-1'), 7.60 (1H, s, H-6) and 9.95 (1H, br. s, NH);  $\delta_C$ ( $CDCl_3$ ) 38.3 ( $CH_2$ -7), 42.1 ( $CH_2$ -2'), 64.0 ( $CH_2$ -5'), 74.5 (CH-4'), 83.2 (CH-3'), 86.0 (CH-1'), 109.4 (C-5), 135.4 (qCN), 137.5 (CH-6), 150.3 (CO), 162.2 (CO), 170.7 and 170.8 ( $OCH_3$ -5' and  $OCH_3$ -3');  $m/z$  ( $CI^+$ ) 383.0785( $M^+$ )  $C_{15}H_{17}N_3O_7S$  requires 383.0792;  $m/z$  (%) 383.1 (90), 326.1 (100), 324.1 (85) and 309.0 (60).

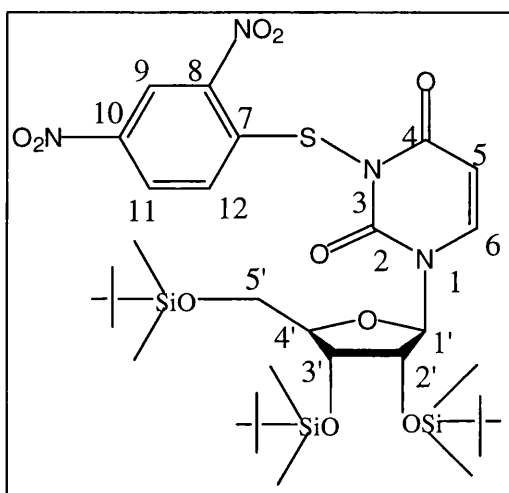
**Bis-[3',5'-O-Diacetyl-2'-deoxyuridine-5 methyl]sulfide SH P22**



To the bromide **SH P19** (1 g, 2.5 mmol) in dimethyl formamide (20 ml) was added sodium hydrosulfide (332 mg, 5.9 mmol). The solution was heated, with stirring, to 70 °C and left overnight. The reaction was diluted with  $H_2O$  (50 ml) and extracted with chloroform ( $2 \times 50$  ml). The organic layer was washed with  $H_2O$  (1 litre) to remove the dimethylformamide, dried with  $MgSO_4$ , filtered and concentrated *in vacuo* yielding crude sulfide (500 mg, 28.4%). The crude sulfide (263 mg, 0.38 mmol) was further purified by column chromatography on silica eluting with ethyl acetate, yielding the title compound as white flakes (180 mg, 68.4%).

New data:  $R_f = 0.26$  (ethyl acetate);  $\nu_{\max}$  (KBr disc)/ $\text{cm}^{-1}$  3436m (NH), 2925w (CH), 2365s (SH), 1743s (CO) and 1685s (C=C);  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) 2.05 and 2.06 (12H,  $2 \times \text{s}$ ,  $2 \times \text{OCH}_3\text{-5'}$  and  $2 \times \text{OCH}_3\text{-3'}$ ); 2.28 (2H, m,  $2 \times \text{H-2'}$ ), 2.40 (1H, ddd,  $J = 14.2, 5.2$  and  $1.8$  Hz,  $2 \times \text{H-2''}$ ), 3.39 (4H, dd,  $J = 14$  and  $14$ ,  $2 \times \text{CH}_2\text{-7}$ ), 4.30 (6H, m,  $2 \times \text{H-5'}$ ,  $\text{H-5''}$  and  $\text{H-4'}$ ), 5.18 (2H, m,  $2 \times \text{H-3'}$ ), 6.26 (2H, dd,  $J = 5.6$  and  $5.6$ ,  $2 \times \text{H1'}$ ) and 7.19 (2H, s,  $2 \times \text{H-6}$ );  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ), 21.3 ( $2 \times \text{OCH}_3\text{-5'}$  and  $2 \times \text{OCH}_3\text{-3'}$ ), 27.2 ( $2 \times \text{CH}_2\text{-7}$ ), 37.7 ( $2 \times \text{CH}_2\text{-2'}$ ), 64.2 ( $2 \times \text{CH}_2\text{-5'}$ ), 74.6 ( $2 \times \text{CH-4'}$ ), 82.6 ( $2 \times \text{CH-3'}$ ), 85.6 ( $2 \times \text{CH-1'}$ ), 112.8 ( $2 \times \text{C-5}$ ), 137.9 ( $2 \times \text{CH-6}$ ), 150.3 (CO), 163.3 (CO), 170.7 and 171.0 ( $2 \times \text{OCH}_3\text{-5'}$  and  $2 \times \text{OCH}_3\text{-3'}$ );  $m/z$  (FAB)  $(\text{M}+\text{H})^+$  682.1864  $\text{C}_{28}\text{H}_{35}\text{N}_4\text{O}_{14}\text{S}$  requires 682.1873;  $m/z$  (%) 683.0 (35), 449.3 (10), 325.0 (30) and 201.0 (100).

2',3',5'-Bis -*O*-(*t*-butyldimethylsilyl)-[ $\text{N}^3$ -(2,4-dinitrophenyl)thio]uridine **SH P23**



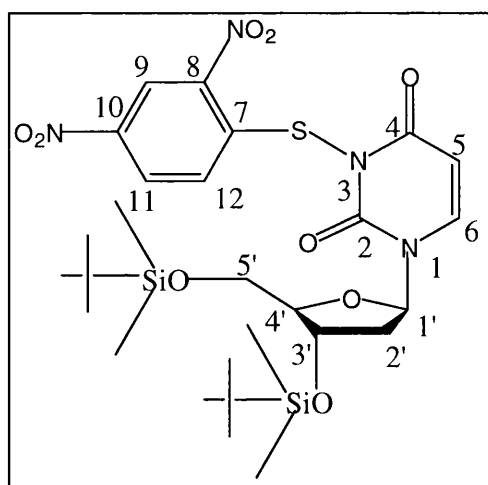
To **SH P6** (100 mg, 0.17 mmol) in dry THF (5 ml) *N, N, N', N'*-tetramethylethylenediamine (TEMED; 150  $\mu\text{l}$ , 1.02 mmol) was added, and the mixture was cooled to  $-78^\circ\text{C}$  using a dry ice-acetone bath. After the addition of *sec*-BuLi (0.78 ml, 1.02 mmol) the mixture was stirred for a further 30 min. A solution of 2, 4-dinitrobenzene sulfenyl chloride (280 mg, 1.19 mmol) in dry THF (3 ml) was added. The reaction mixture was poured into  $\text{H}_2\text{O}$  (50 ml), and extracted with ethyl acetate ( $3 \times 50$  ml). The organic layer was separated and



washed with brine (10 ml), dried with  $\text{MgSO}_4$  and concentrated *in vacuo*, yielding the crude product which was purified using column chromatography on silica, eluting with 2:1 ether: hexane, yielding the product as a yellow solid (120 mg, 50.9%).

New data:  $R_f = 0.73$  (2:1 ether: hexane);  $\delta_H$  ( $\text{CDCl}_3$ ) -0.01, -0.02, 0.00 and 0.01 (18H,  $6 \times s$ ,  $6 \times \text{SiMe}$ ), 0.80 and 0.85 (27H,  $3 \times s$ ,  $3 \times t$ -butyl), 3.69 (1H, d,  $J = 11.5$ , H-5'), 3.87 (1H, d,  $J = 11.5$ , H-5''), 3.94-4.13 (3H, m, H-2', H-3' and H-4'), 5.90 (1H, d,  $J = 5.5$ , H-5), 6.04 (1H, d,  $J = 6.9$ , H-1'), 7.94 (1H, d,  $J = 5.3$ , H-6), 7.92 (1H, d,  $J = 9.1$ , H-8), 7.94 (1H, d,  $J = 5.3$ , H-6), 8.17 (1H, dd,  $J = 9.9$  and 2.4, H-9) and 8.95 (1H, d,  $J = 2.4$ , H-11);  $\delta_C$  ( $\text{CDCl}_3$ ) -5.2 to -3.8 ( $6 \times \text{SiMe}$ ), 18.3, 18.4 and 18.8 ( $3 \times \text{qC}$ ), 26.1, 26.2 and 26.3 ( $3 \times \tau$ -butyl  $\text{CH}_3$ ), 62.1 (C-5), 71.0 (C-4), 76.5 (C-2), 84.8 (C-3), 89.3 (C-1), 102.2 (C-5), 120.6 (C-8), 126.2 (C-9), 127.86 (C-11), 136.0 (C-7), 138.1 (C-10), 138.5 (C-12), 140.7 (C-6), 150.6 (CO) and 163.7 (CO);  $m/z$  (FAB) 807.2927 ( $\text{M} + \text{Na}$ ) $^+$   $\text{C}_{33}\text{H}_{56}\text{N}_4\text{O}_{10}\text{Si}_3\text{Na}$  requires 807.2921  $m/z$  (%) 807.2 (48), 610.6 (100), 366.4 (55).

### 3',5'-Bis -*O*-(*t*-butyldimethylsilyl)-[ $\text{N}^3$ -(2,4-dinitrophenyl)thio]-2'-deoxyuridine SH P24

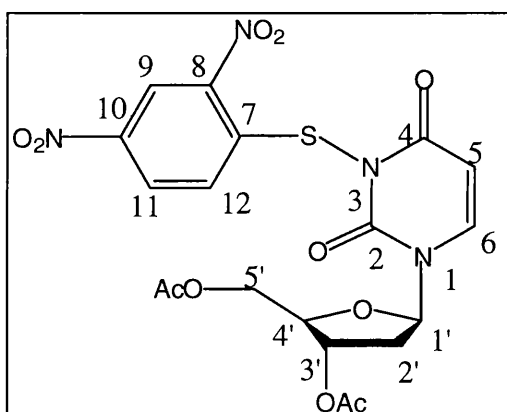


To **SH P5** (80 mg, 0.17 mmol) in dry THF (4 ml) was added *N, N, N', N'*-tetramethylethylenediamine (TEMED; 157  $\mu\text{l}$ , 1.05 mmol) was added, and the mixture was cooled to  $-78^\circ\text{C}$  using a dry ice-acetone bath. After the addition of

*sec*-BuLi (0.80 ml, 1.05 mmol) the mixture was stirred for a further 30 min. A solution of 2, 4-dinitrobenzene sulfonyl chloride (286 mg, 1.22 mmol) in dry THF (3 ml) was added. The reaction mixture was poured into H<sub>2</sub>O (50 ml), extracted with ethyl acetate (3 × 50 ml). The organic layer was separated and washed with brine (10 ml), dried with MgSO<sub>4</sub> and concentrated *in vacuo*, yielding the crude product which was purified using column chromatography on silica, eluting with 2:1 ether: hexane, yielding the product as a yellow solid (20 mg, 17.4%).

New data:  $R_f = 0.41$  (2:1 ether: hexane);  $\delta_H$  (CDCl<sub>3</sub>) -0.03, -0.02, 0.00 and 0.01 (12H, 4 × s, 4 × SiMe), 0.78 and 0.81 (18H, 2 × s, 2 × *t*-butyl), 1.95 (1H, m, H-2'), 2.22 (1H, m, H-2''), 3.64 (1H, m, H-4'), 3.81 (3H, m, H-4' and H-5'), 4.31 (1H, m, H-3'), 5.57 (1H, d,  $J = 8.1$  Hz, H-5), 6.18 (1H, m, H-1'), 7.79 (1H, d,  $J = 8.1$ );  $m/z$  (FAB) 674.8585 ( $M + Na$ )<sup>+</sup> C<sub>27</sub>H<sub>39</sub>N<sub>4</sub>O<sub>9</sub>Si<sub>2</sub>Na requires 674.8599.

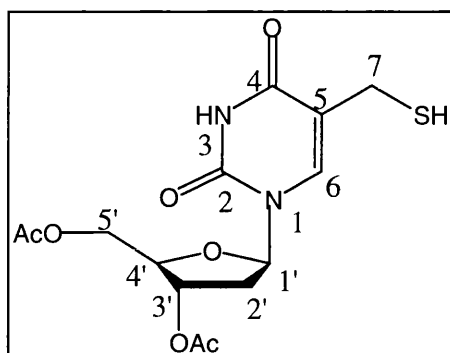
3',5'-*O*-Diacetyl-5-[N<sup>3</sup>-(2,4-dinitrophenyl)thio]-2'-deoxyuridine **SH P25**



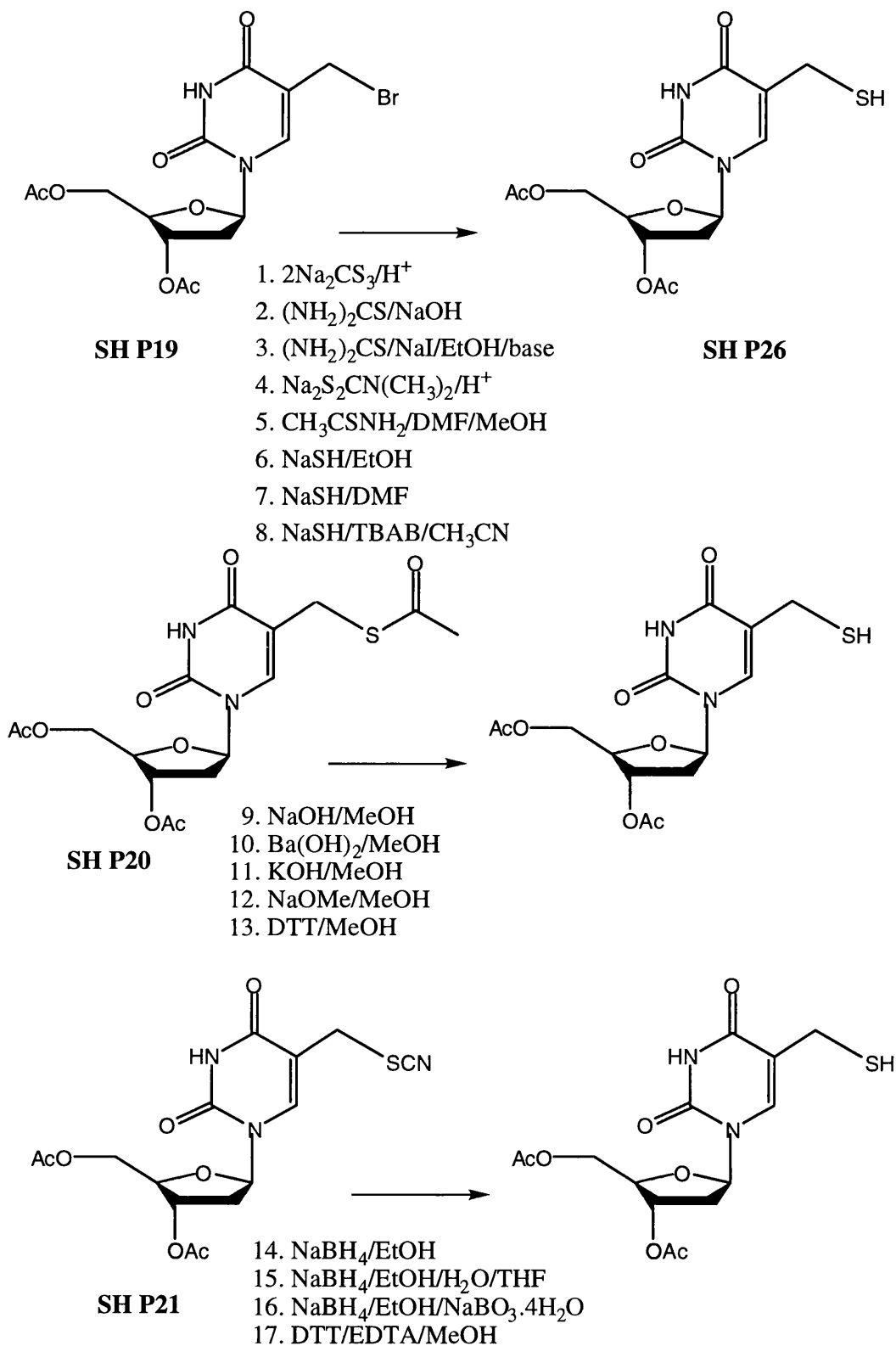
3',5'-*O*-Diacetyl-2'-deoxyuridine **SH P9** (156 mg, 0.50 mmol) and silver trifluoroacetate (166 mg, 0.75 mmol) in anhydrous MeCN (4 ml) were stirred at room temperature. To this was added 2, 4-dinitrobenzenesulfonyl chloride (175 mg, 0.75 mmol) and then the reaction was stirred for a further 2 h. Tlc analysis showed a new product had formed. The reaction mixture was filtered to removed the AgCl and the filtrate was concentrated *in vacuo* yielding the crude product. This was further purified by column chromatography on silica, eluting with 9:1 dichloromethane: acetone, yielding (10 mg, 5%) of product;  $\delta_H$  (CDCl<sub>3</sub>) 2.04 and

2.07 (3H, 2 × s, OCH<sub>3</sub>-5' and OCH<sub>3</sub>-3'), 2.19 (1H, m, H-2'), 2.53 (1H, ddd, J = 14.3, 5.8 and 2.5, H-2''), 4.23-4.27 (3H, m, H-4' and 2 × H-5'), 5.17 (1H, dd, J = 6.3, H-3'), 5.93 (1H, d, J = 8.2, H-5), 6.19 (1H, dd, J = 6.6, H-1'), 7.05 (1H, d, J = 9.0, H-7), 7.27 (1H, d, J = 9.8, H-8), 7.61 (1H, d, J = 8.2, H-6), 8.31 (1H, dd, J = 9.0 and 2.4, H-9) and 9.09 (1H, d, J = 2.4, H-10); *m/z* (CI<sup>+</sup>) 509.4235  
 C<sub>19</sub>H<sub>17</sub>N<sub>4</sub>O<sub>11</sub>S requires 509.4240; *m/z* (%) 509.4 (20), 326.1 (90), 201.1 (20) and 81.0 (100).

## 2.9 Attempted Synthesis of 3',5'-*O*-Diacetyl-5-thiomethyl-2'-deoxyuridine SH P26



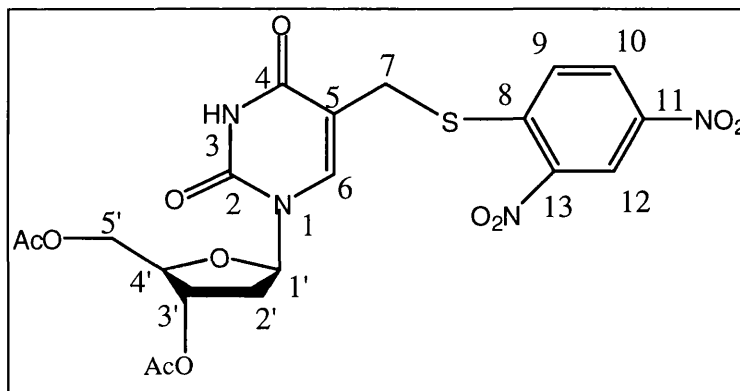
The brominated derivative **SH P19** was reacted with a variety of reagents to convert it into the mercaptomethyl derivative **SH P26**. These included reactions with sodium hydrosulfide (Langen and Barwolff, 1975), sodium trithiocarbamate (Martin and Greco, 1968), thiourea (Urquhart *et al.*, 1955) sodium iodide and thiourea (Melnick and Weinreb, 1988), *N,N*-dimethyldithiocarbamate (Kulka, 1956) and thioacetimide (Giner-Sorolla and Medrek, 1966). Several methods were attempted to hydrolyse 3',5'-*O*-diacetyl-5-thioacetylmethyl-2'-deoxyuridine **SH P20**. Several methods of reducing 3',5'-*O*-diacetyl-5-thiocyanatomethyl-2'-deoxyuridine **SH 21** were attempted. These included reduction with sodium borohydride (Hideg *et al.*, 1992; Marquet *et al.*, 1993; Campanini *et al.*, 1995) and DTT (Hanna and Meyer, 1996). The results of these reactions will be discussed in Section 2.11 and are illustrated in Figure 2.12.



**Figure 2.12 Attempted syntheses 3',5'-O-Diacetyl-5-thiomethyl-2'-deoxyuridine SH P26**

Diagram summarising the attempted syntheses of the 5-mercaptopmethyl-2'- deoxyuridine **SH P26**.

## 2.10 Attempted Synthesis of 3',5'-O-Diacetyl-5-[S-2,4-dinitrophenyl]-thiomethyl-2'-deoxyuridine SH P27

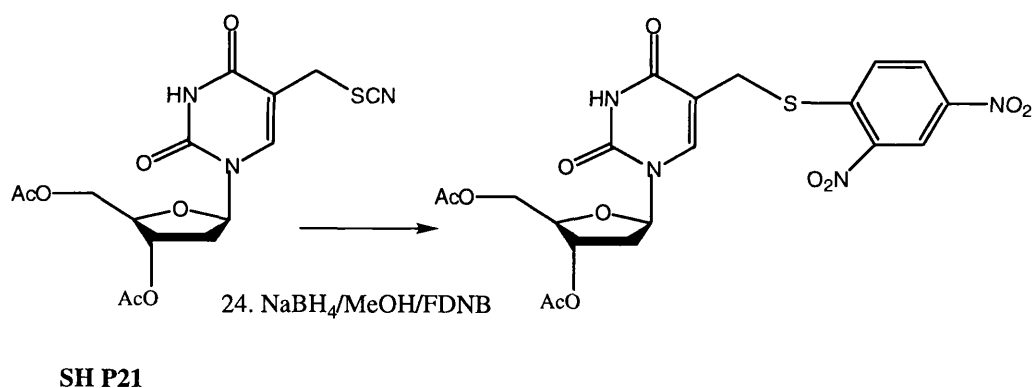
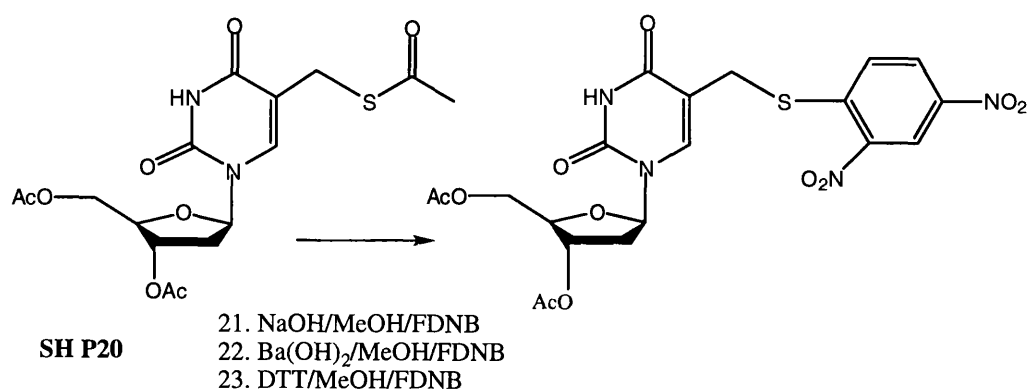
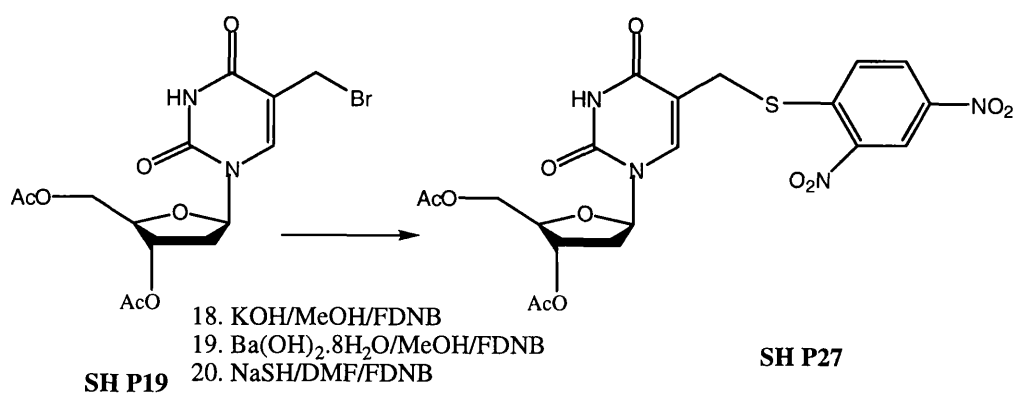


Several methods were used to generate 3',5'-O-diacetyl-5-thiomethyl-2'-deoxyuridine **SH P27**. These included the methods to generate the free thiol (Section 2.9), and attempts to enact *in situ* protection of the thiol **SH P26**, with 2,4-dinitrobenzene sulfenyl chloride. These reactions are illustrated in Figure 2.13 and will be discussed in Section 2.11.

## 2.11 Discussion

Prior to the synthesis of the diazirine-containing phosphoramidite nucleotide, a simple diazirine was synthesised to ascertain the feasibility of the reaction. The diazirine of laevulinic acid was prepared in three steps with an overall yield of 14.6%. Another intermediate that was required for the synthesis of the diazirine target contained an aldehyde group at the 5-position of thymidine. This involved the synthesis of a hydroxymethyl precursor. The synthesis of this compound proved to be very poor, with yields of 8% or less. The diazirine target was therefore omitted from this study. Diazirines are often stabilised by having an adjacent trifluoroacetate group (Lui, 1986). For the above reasons, the synthesis of the diazirine target was omitted from this study.

The disulfide of 2, 4-dinitrobenzene was obtained in 84.3% yield. The reduction of the disulfide bond yielded the monomer thiol (Kharasch *et al*, 1955). The



**Figure 2.13 Attempted synthesis of 3',5'-O-Diacetyl-5-[S-2,4-dinitrophenyl]thiomethyl-2'-deoxyuridine SH P27**

Illustration of the direct methods used to synthesise the 3',5'-O-Diacetyl-5-thiomethyl-2'-deoxyuridine (section 2.11).

reduction used glucose/NaOH (Bordwell and Andersen, 1953). However, this reaction proceeded in very low yield (less than 10%). The disulfide was not soluble in methanol or benzene, the two solvents in which the palladium cross coupling reaction is normally performed (Bergstrom *et al.*, 1991). It was also insoluble in THF, the solvent required for the lithiation reaction (Armstrong *et al.*, 1989).

The lithiation reaction was carried out on both uridine and 2'-deoxyuridine to allow for optimisation of reaction conditions. The starting nucleoside 2',3',5'-bis-O-[ *t*-butyldimethylsilyl)dimethylsilyl]-uridine **SH P5** and 3',5'-bis-O-[ *t*-butyldimethylsilyl)dimethylsilyl]-2'-deoxyuridine **SH P6** were obtained in yields of 93.0% and 95%, respectively using TBDMSiCl, DMF, and imidazole as catalyst at 50 °C (Ogilvie *et al.* 1978). For the lithiation reactions the nucleoside was dissolved in dry THF and the required amount of sec-BuLi was added, in the presence of TEMED. The lithium salts were quenched with 2,4-dinitrobenzenesulfenyl chloride.

The methods for direct attachment to the pyrimidine ring with the 2, 4-dinitrobenzene-protected thiol gave mixture of both N-3 and C-5 lithiated products. Lithiation of **SH P5** and **SHP6**, followed by quenching with 2,4-dinitrobenzenesulfenyl chloride, resulted in N-3 arylated products, indicating that the hydrogen at N-3 is more acidic in the pyrimidine ring than the C-5 position under the reaction conditions tried. **SH P23** is contaminated with the 2',3',5'-Bis - *O*-(*t*-butyldimethylsilyl)-[C-5,(2,4-dinitrophenyl)thio]uridine as indicated by the C-5 lithiated product at (*m/z*; (FAB) 808.2960 C<sub>33</sub>H<sub>57</sub>N<sub>4</sub>O<sub>10</sub>Si<sub>3</sub>Na requires 808.3110) at 27.3% intensity compared to 47.6 % for the N-3 lithiated species. In addition there was another major species at (*m/z*; (FAB) 809.2964 C<sub>33</sub>H<sub>58</sub>N<sub>4</sub>O<sub>10</sub>Si<sub>3</sub>Na requires 809.3107) at 16.6% intensity. Similarly **SH P24** was contaminated with N-3 species that co-eluted with the C-5 lithiated species during column chromatography. Similarly, the reaction using silver reagents generated exclusively N-3 thioarylated product **SH P25**.

The method used for incorporation of the thiol group required thiocyanation of the pyrimidine ring. Initial reported yields of this reaction on 2'-deoxyuridine were very low (10%), although there are literature reports (Bradley and Hanna, 1992) of up to 54% for the reaction on the unprotected form. The hydroxyls of 2'-deoxyuridine were protected as acetates in **SH P9** yields ranging from 80-82%. Protection of the sugar hydroxyls by acetylation prior to thiocyanation resulted in higher yields (Nagamachi *et al.*, 1974) of up to 95%. The thiocyanation reaction involves generation of a nucleophilic 'Cl-SCN' species *in situ* by adding chlorine gas to a solution of dry glacial acetic acid. Prior to addition of the potassium thiocyanate, the amount of chlorine added was determined by iodometric techniques (Flaschka *et al.*, 1969). The reaction proceeds by nucleophilic substitution of the pyrimidine ring on the 'Cl-SCN'. The thiocyanation went well yielding 5-thiocyanato-2'-deoxyuridine **SH P10** in 73.5% yield.

Reduction of the thiocyanato group was achieved under mild reaction conditions using dithiothreitol, which displaced the cyano group yielding the mercaptan **SH P11** in 75.3% yield. Protection of the thiol was carried out using 2,4-dinitrofluorobenzene, in dry acetonitrile with triethylamine as a catalyst, yielding the protected thiol **SH P12** in 62.8%. The deacetylation reaction *via* methanolysis of the diacetate with sodium methoxide in methanol gave the dihydroxyl **SH P13** in yields of 86.3%.

Protection of the 5'-hydroxyl as the dimethoxytrityl ether was carried out using two methods: the direct use of dimethoxytrityl chloride in pyridine; and using the tetrafluoroborate salt of dimethoxytrityl in pyridine. Hanna and Meyer (1996) suggested that the use of 4, 4'-dimethoxytrityl chloride in pyridine was an unsuitable method for protection of the 5'-hydroxyl group, and found increased yields using a method that used an activated trityl salt (Bleasdale *et al.*, 1989). In contrast to their results it was found that the reverse was true. The trityl chloride method was superior, yielding the trityl-protected derivative **SH P14** in yields of 84.0%.



It was desirable to make two different thiol-containing analogues, one with the thiol directly attached to the C-5 position and a second with a mercaptomethyl at C-5. The starting material for the 5-mercaptomethyl analogue was thymidine.

Protection of the hydroxyl groups using acetic anhydride in pyridine, as before, gave high yields of the diacetate **SH P18** (80.8%). A precursor to the mercaptomethyl compound was the bromomethyl derivative **SH P19**. The method for photochemical bromination (Sergiev *et al.* 1997) using bromine in dichloroethane (DCE) was applied to the 3', 5'-diacetyl-thymidine **SH P18**, and gave the brominated derivative in 44.4% yield. Another method, involving bromination with *N*-bromosuccinimide in carbon tetrachloride (Shealy *et al.*, 1986) did not yield any of the desired product, due to poor solubility of the starting material. The bromide was reacted with potassium thioacetate in dimethylformamide; the thioacetatomethyl derivative **SH P20** was made in 53.7% yield. In an analogous reaction, (Langen and Barwolff, 1975) the bromide reacted with potassium thiocyanate in dimethylformamide occurred with S<sub>N</sub>2 displacement; the thiocyanatomethyl **SH P21** (Langen and Barwolff, 1975) was obtained in 58.6% yield.

The methods used to attempt to synthesise 3',5'-*O*-diacetyl-5-thiomethyl-2'-deoxyuridine **SH P26** are summarised in Figure 2.12. The reactions of the bromomethyl derivative illustrated in Figure 2.12.1-2.12.5 were carried out according to the literature methods (2.12.1; Martin and Greco, 1968); (2.12.2; Urquhart *et al.*, 1955); (2.12.3; Melnick and Weinreb, 1988); (2.12.4; Kulka, 1956); (2.12.5; Giner-Sorolla and Medrek, 1966). However, the only products from these reactions were unreacted starting material or decomposition products. However, reaction 2.12.6 and 2.12.7 (Baerwolff and Langen, 1975), resulted in sulfide **SH P22**. The reactions that involved the selective hydrolysis of the thioacetyl moiety of **SH P20** (Figure 2.12.9-12) used aqueous methanolic hydrolysis using 0.5-1 equivalent of base. Under these conditions as well as hydrolysis of the thioacetyl group, hydrolysis of the hydroxyl acetyl protecting

groups also occurred yielding a species that was highly charged and difficult to extract using organic solvents, making further purification difficult. Sodium borohydride reductions of the thiocyanatomethyl derivative **SH P21** (Figure 2.12.14-16) did produce the thiol moiety, but subsequent aqueous work up resulted in the sulfide **SH P22** as the predominant product.

The failure to synthesise 3',5'-*O*-diacetyl-5-thiomethyl-2'-deoxyuridine **SH P26** was due to a number of factors. The extreme reactivity of the mercaptomethyl moiety generated during the reactions mentioned above was probably due to the proximity of the pyrimidine ring. The delocalisation of the electrons in the ring would push electrons towards the sulfur atom, thus increasing its nucleophilicity, resulting in the observed kinetic product, the sulfide **SH P22**. Any mercaptomethyl group that was formed immediately reacted with another molecule, forming the sulfide with the release of H<sub>2</sub>S.

The reaction of the bromomethyl compound with sodium hydrosulfide predominantly yielded the sulfide **SH P22**, usually contaminated with other thiol species. The high resolution mass spectra of the sulfide, that had been purified by column chromatography, showed in addition to the peak for the sulfide ( $m/z$  682.1), a peak at  $m/z$  715.8 corresponding to the disulfide (C<sub>28</sub>H<sub>35</sub>N<sub>4</sub>O<sub>14</sub>S<sub>2</sub> requires 715.1590; found 715.1592) was also present. The peak in the IR spectrum in the region where free thiols absorb indicates that some free thiol is present. Thus, the sulfide was contaminated with both the disulfide and the free thiol. Thin layer chromatography of the crude reaction mixture showed spots corresponding to the bromide compound **SH P19**, the diacetylated thymidine **SH P9**, and slower moving species below R<sub>f</sub> = 0.3 (ethyl acetate). These slower moving products corresponded to the three different thiol species that co-eluted during column chromatography.

The reactions 18-21 in Figure 2.13 were carried out in order that any free thiol generated, either by hydrolysis or reduction of **SH P19**, **SH P20** and **SH P21**

could be protected *in situ*, by the addition of 2,4-dinitrobenzenesulfonyl chloride to the reaction mixture, yielding 3',5'-*O*-diacetyl-5-[2,4-dinitrophenyl]thiomethyl-2'-deoxyuridine **SH P27**. Products were obtained with reactions 2.13.19, 2.13.21 and 2.13.23. These had a  $m/z$  value of 524.6, corresponding to the desired product but in very low concentration. The  $^1\text{H}$  NMR spectra show product signals, but more than one set of aromatic signals are present, indicating that the product is contaminated with aromatic species that co-elute during chromatography.

In total three methods for making the phosphoramidites were used. The first used tetrazole and 2-cyanoethyl-*N, N, N', N'*-tetraisopropylphosphorodiamidite (Method A), the second used tetrazole diisopropylammonium salt and 2-cyanoethyl-*N, N, N', N'*-tetraisopropylphosphorodiamidite (Method B), and the third involved the use of the chlorophosphine reagent (Method C) as the phosphitylating reagent.

The diisopropyl ammonium salt of tetrazole (D. Picken, pers. commun.) was used to prevent any possibility of overactivation of the phosphoramidite, which could lead to polymerisation; i.e. the addition of an extra alcohol monomer to the activated trivalent phosphorus atom. Of the three methods used, the method that utilised the diisopropyl ammonium salt of tetrazole to activate the phosphitylating reagent was the favoured one (Method B). This method involving two precipitations of the crude phosphoramidite, followed by column chromatography with ethyl acetate as the eluant gave yields of over 50%. Method A involving a chromatography step with methanol and chloroform and the eluants, tended to cause oxidation of the phosphoramidite and lower yields of around 30%. Extra peaks in the  $^{32}\text{P}$  NMR spectra of the product made by method A showed that both oxidation products and dimerisation of the phosphoramidite had occurred. Method C, although it gave pure products, was rather low-yielding; on average 28% phosphoramidite was obtained. Due to the air sensitivity of the phosphoramidite it was impossible to obtain elemental analysis by combustion microanalysis.

The  $^1\text{H}$  NMR spectrum of **SH P17** is shown in Figure 2.14.

## 2.12 Conclusions

The direct methods for attaching the 2,4-dinitrophenyl protected thiol moiety to the C-5 position of the pyrimidine ring were unsuccessful. As discussed above the reactions gave the wrong products, or the yields were too low to be of synthetic use. This meant that the protected thiol had to be made by indirect methods.

5'-*O*-(Dimethoxytrityl)-5-[*S*-(2,4-dinitrophenyl)-thio]-2'-deoxyuridine-3'-*O*-(2-*N,N'*-cyanoethyl diisopropylphosphoramidite **SH P17** was synthesised.

Unfortunately, synthesis of the 5-mercaptomethyl phosphoramidite analogue proved too problematic, and only side products were isolated. The C-5 mercaptomethyl derivative is more reactive than the C-5 thiol, which is deactivated by the pyrimidine ring.

The phosphoramidite **SH P17** was incorporated into oligonucleotides and tested as a crosslinking agents for Tn3 resolvase (Chapter 5).

Another approach that could have been tried to generate the mercaptomethyl derivative would have been to use methodology of Gupta *et al.*, (1975). This involved the initial synthesis of the mercaptomethyl uracil, followed by coupling to the sugar group and isolation of the desired anomer. Although separation of the anomers can be difficult following this type of coupling reaction, it is still a feasible route to the mercaptomethyl derivative, and given the time over again this route could have been done in parallel alongside the other methods tried.

With more time an alternative thiol-containing pyrimidine nucleotide could have been synthesised. This analogue contains a thiol group at the end of a two carbon linker (Figure 2.8, compound 3). The synthesis of this analogue would have provided a means of comparing the reactivity of the thiol groups at the C-5 position when they are directly attached to the pyrimidine ring system, or

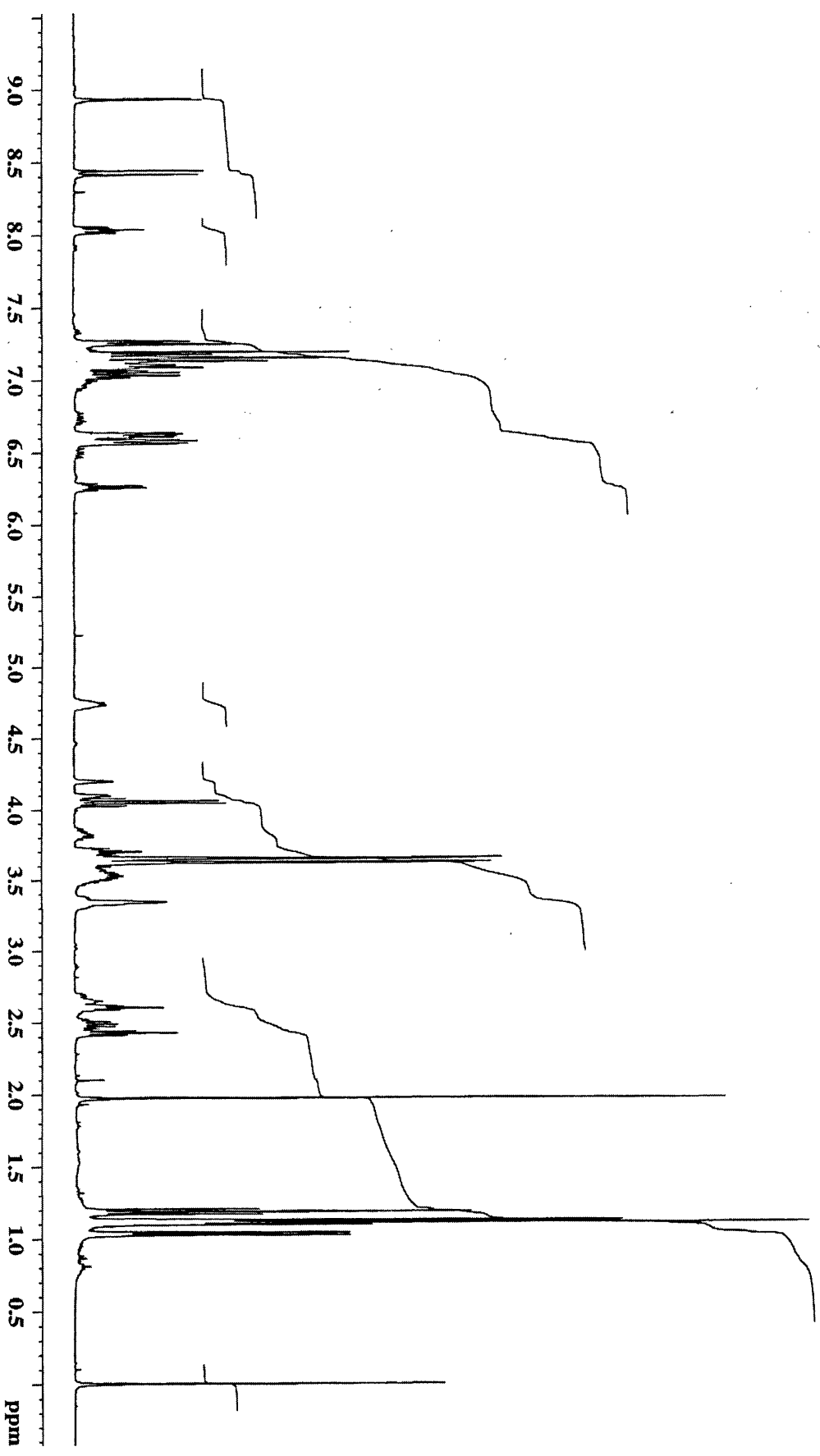


Figure 2.14  $^1\text{H}$  spectrum of 5'-O-(Dimethoxytrityl)-5-[S-(2,4-dinitrophenyl)-thio]-2'-deoxyuridine-3'-O-(2-N,N''-cyanoethyl)diisopropylphosphoramidite SH P17

separated from the ring by one or two carbon chain linkers. The synthesis of this compound has been carried out using two different methods (Goodwin and Glick, 1993); in total, seven synthetic steps are required to produce the required protected phosphoramidite derivative. On reflection, this synthesis could have been carried out in parallel with the synthesis attempted in this work. It is possible to predict that the thiol that was separated from the pyrimidine ring by a two carbon chain would be less influenced by the ring, and should be stable to nucleophilic displacement leading to loss of sulfur.

## **Chapter 3: Materials and methods**

### 3.1 Bacterial Strains

During the project the following bacterial strains were used, as illustrated in Table 3.1.

**Table 3.1**

Strain	Genotype	Source
AB1157	<i>thr1, leu6, hisG4, thi1, ara14, proA2, argE3, galK2, sup37, xyl15, mtl1, tsx33, str31</i>	
DS941	AB1157, but <i>recF143, supE44, lacZΔM15, lacI<sup>q</sup></i>	D. J. Sherratt
DS902	AB1157, but <i>recA13</i>	D. J. Sherratt
DH5α	<i>supE44, ΔlacU169(ΦlacZΔM15)hsdR17, recA1, endA1, gyrA96, thi<sup>-1</sup>, relA1</i>	Gibco BRL
BL21	<i>E.coli B, F<sup>-</sup>, dcm, ompT, hsdS(r<sub>B</sub>m<sub>b</sub><sup>-</sup>)gal</i>	W. Studier, 1990
BL21(DE 3)	<i>hsd, gal, (λcI ts857, ind1, Sam7, ini5, lac<sub>UV5</sub>-T7 gene-1)</i>	W.Studier, 1990
BL21(DE 3) pLysS	As BL21(DE3), but has the T7 lysozyme expression plasmid pLysS	W. Studier, 1990

### 3.2 Plasmids

The plasmids that were used and/or constructed for the work are listed in Table 3.2.

### 3.3 Chemical Reagents

The chemicals used in the molecular biology and the organic synthesis work are listed in Table 3.3.



Table 3.2 Plasmids

Plasmids	Size (bp)	Antibiotic Marker	Description/Derivation	Source/reference
PUC18	2686	Ap	Standard Cloning Vector	Yanisch -Perron <i>et al.</i> , 1985
pMTL23	2505	Ap	Cloning vector	CAMR, Porton Down
pBR322	4361	Ap.Tc	Standard Cloning Vector	Sutcliffe, 1978
pUC71K	3966	Ap Km	Cloning vector, source of Km <sup>r</sup> gene	Taylor and Rose, 1982
POG5	2639	Ap	130 bp <i>Sst</i> I/ <i>Xba</i> I RI <i>res</i> + 2505 bp <i>Sst</i> I/ <i>Xba</i> I pMTL23	Blake <i>et al.</i> , 1995 (figure 4.16)
POG3	4111	Ap Km	Derivative of POG5 and pUC71K, with two RI <i>res</i> sites in direct repeat	Mellwraith <i>et al.</i> , 1996
pMM5	4903	Ap.Tc	2623 bp POG5 <i>Hind</i> III- <i>Hinc</i> II + 2280 bp pMA21 <i>Eco</i> RI/ <i>Hind</i> III digest	Mellwraith <i>et al.</i> , 1996
pDB34	8500	Km	3060 bp <i>Nde</i> I/ <i>Pst</i> I (partial) Km pDB703 + 5400 bp <i>Nde</i> I/ <i>Pst</i> I pDB4	Blake, 1993
pMS53	4903	Ap.Tc	derivative of pMM5, one <i>res</i> site GATATC	M. Burke
pMSH1	2690	Ap	<i>Spe</i> I/ <i>Bam</i> HI digest of pMS53 resolved + <i>Eco</i> RV digest of pMM5	Section 6.2
pMSH2	3255	Ap	pMSH1 <i>Hind</i> III + 565 bp <i>Hind</i> III digest	Section 6.2
pSA1101	6692	Km	Tn3+ resolvase expression vector	S. J. Rowland; Arnold <i>et al.</i> , 1999
pALY15	6693	Km	pSA1101 containing T162C mutation.	MacDonald, 1999

pAT5	6240	Ap	resolvase open reading frame with the first two-thirds subdivided into clonable fragments in pSELECT.	Arnold <i>et al.</i> , 1999
ApAT5	5465	Ap	pAT5 with <i>EcoRV-NruI</i> fragment deleted	M. Boocock
ApAT6	5464	Ap	ApAT5 with extra restriction sites between <i>HindIII-AgeI</i>	M. Boocock
pMA61	6692	Ap.Tc	Tn3+ expression vector with pAT5 resolvase open reading frame at <i>NdeI-Asp718</i>	M. Boocock
pMA6211	5466	Ap.Tc	ApAT5 + <i>EagI-Asp718</i> fragment (T162C) from pALY15 with extra restriction sites in the C-terminal domain.	M.Boocock
pMS81	5466	Ap.Tc	pMA6211 that has <i>HpaI-BstHI</i> replaced with oligonucleotides that introduce new restriction sites (and remove T162C)	M.Burke, 2000
pMS82	4948	Ap	from <i>PstI</i> (partial) and <i>SspI</i> digest of pMS81. Mutant oligonucleotides will be cloned into this plamid	M.Burke, 2000
pMS88	4948	Ap	<i>MfeI-PstI</i> pMS82 replaced with oligonucleotides containing the A171C mutation	M.Burke, 2000
pMS89	4948	Ap	<i>MfeI-PstI</i> pMS82 replaced with oligonucleotides containing the R172C mutation	M.Burke, 2000
pMS90	4948	Ap	<i>MfeI-PstI</i> pMS82 replaced with oligonucleotides containing the S173C mutation	M.Burke, 2000
pMS91	4948	Ap	<i>MfeI-PstI</i> pMS82 replaced with oligonucleotides containing the Y176C mutation	M.Burke, 2000
pMS92	4948	Ap	<i>PstI-SalI</i> pMS82 replaced with oligonucleotides containing the K177C mutation	M.Burke, 2000
pMS100	5466	Ap.Tc	pMS82 <i>EagI-Asp718</i> + ApAT6 <i>EagI-Asp718</i> , fully cassetted resolvase reading frame	M.Burke, 2000 Figure 5.11
PMS95	6693	Km	pMS88 <i>EagI-Asp718</i> 145 bp + 6548 bp <i>EagI-Asp718</i> pALY15, over-expression vector for A171C mutation	Section 4.5

PMS96	6693	Km	pMS89 <i>EagI-Asp</i> 718 145 bp + 6548 bp <i>EagI-Asp</i> 718 pAL Y15 over-expression vector for R172C mutation	Section 4.5
PMS97	6693	Km	pMS90 <i>EagI-Asp</i> 718 145 bp + 6548 bp <i>EagI-Asp</i> 718 pAL Y15, over-expression vector for S173C mutation	Section 4.5
PMS98	6693	Km	pMS91 <i>EagI-Asp</i> 718 145 bp + 6548 bp <i>EagI-Asp</i> 718 pAL Y15, over-expression vector for Y176C mutation	Section 4.5
pMS99	6693	Km	pMS92 <i>EagI-Asp</i> 718 + 6548 bp <i>EagI-Asp</i> 718 pAL Y15, expression vector for K177C mutation	Section 4.5
pMS20	3326	ApTc	From pMA2350 by mutating both <i>res</i> to 'AC' at centre	Mellwraith <i>et al.</i> , 1997
pMSHAres	3325	Ap	pMS20 cut <i>EcoRI-HindIII</i> filled + pMM5 cut <i>BamHI</i>	S. Hardie, 2001

**Table 3.3**

Chemicals	Source
General chemicals, biochemicals, organic solvents	Aldrich/Sigma, BDH
Media	Difco, Oxoid
Agarose	BRL, FMC
Acrylamide	National diagnostics, BioRad
Radiochemicals	ICN Biochemicals
10 x Restriction enzyme buffers	BRL, New England Biolabs
Reagents for oligonucleotide synthesis	Cruachem
5 x Ligase buffer	BRL

**3.4 Oligonucleotides**

The oligonucleotides used for cloning and for crosslinking are listed in Table 3.4.

**3.5 Antibiotics**

The antibiotics used as selectable markers are listed in Table 3.5.

**Table 3.5**

Antibiotic	Stock Solution	Working concentration
Ampicillin (Ap)	5 mg/ml in dH <sub>2</sub> O	50 µg/ml
Chloramphenicol (Cm)	2.5 mg/ml in ethanol	25 µg/ml
Kanamycin (Km)	5 mg/ml in dH <sub>2</sub> O	50 µg/ml
Tetracycline (Tc)	12.5 mg/ml in 70% ethanol	12.5 µg/ml

**3.6 Bacterial growth media**

For the growth of *E. coli* for DNA preparation, transformations and protein purification.

**L-broth:** 10 g bacto-tryptone, 5 g bacto-yeast extract, 5 g NaCl, made up to 1 litre with deionised water, adjusted to pH 7.5 with NaOH.

**L-agar:** L-broth with 15 g/l agar.

The above media are sterilised at 120 °C for 15 minutes before use.

**Table 3.4 Oligonucleotides**

SHRV 40 nt	5'CACTAGTCTGTTCGAAATATGATATCTTATCAGACATA GG-3'	Section 6.2
SHTT 40 nt	5'CACTAGTCTGTTCGAAATATTATAAATTATCAGACATA GG-3'	Section 3.20
SHBT 48 nt	5'AATTCCTATGTCTGATAATTTATAATTTTCGAACAGAC TAGTGAGCT-3'	Section 3.20
SHR2 48 nt	5'AATTCCTATGTCTGATAATTGATATCATTTCGAACAGA CTAGTGAGCT-3'	Section 6.2
C155A 56 nt	5'AACGCTTCATCAGAAGGGCACTGGTGCAACGGAAATT GCTCATCAATTGAGTATTG 3'	Section 4.5
C155B 60 nt	5'CGCGCAATACTCAATTGATGAGCAATTTCCGTTGCACC AGTGCCCTTCTGATGAAGCGTT 3'	Section 4.5
C155C 23 nt	5'-AATTGAGTATTTGTCGCTCCACGGTTTA-3'	Section 4.5
C155D 24 nt	5'-TAAACCGTGGAGCGACAATACTC-3'	Section 4.5
C155E 28 nt	5'-AATTGAGTATTGCGTGTTCACGGTTTA-3'	Section 4.5
C155F 24 nt	5'-TAAACCGTGGAACACGCAATACTC-3'	Section 4.5
C156A 28 nt	5'-AATTGAGTATTGCACGCTGTACGGTTTA-3'	Section 4.5
C156B 24 nt	5'-TAAACCGTACAGCGTGCAATACTC-3'	Section 4.5
C156C 28 nt	5'-AATTGAGTATTGCACGCTCCACGGTTTG-3'	Section 4.5
C156D 24 nt	5'-CAAACCGTGGAGCGTGCAATACTC-3'	Section 4.5
C156E 32 nt	5'TTGTATTCTTGAAGACGAAAGGGCGAGCTGAGTCGAC- 3'	Section 4.5
C156F 36 nt	5'-TCGACTCAGCTCGCCCTTTCGTCTTCAAGAATACAA-3'	Section 4.5
SHT1 40 nt	5'CACTAGTCTGTTCGAAATATTATAAATTATCAGACATA GG-3'	Section 5.2
SHT3 40 nt	5'CACTAGTCTGTTCGAAATATTATAAATTATCAGACATA GG-3'	Section 5.2
SHT4 40 nt	5'CACTAGTCTGTTCGAAATATTATAAATTATCAGACATA GG-3'	Section 5.2
SHT12 40 nt	5'CACTAGTCTGTTCGAAATATTATAAATTATCAGACATA GG-3'	Section 5.2
SHB1 48 nt	5'AATTCCTATGTCTGATAATTTATAATTTTCGAACAGA CTAGTGAGCT-3'	Section 5.2
SHB3 48 nt	5'AATTCCTATGTCTGATAATTTATAATTTTCGAACAGA CTAGTGAGCT-3'	Section 5.2
SHB5 48 nt	5'AATTCCTATGTCTGATAATTTATAATTTTCGAACAGA CTAGTGAGCT-3'	Section 5.2

SHB13 48 nt	5'AATTCCTATGTCTGATAATTTATAATATTTCGAACAGA CTAGTGAGCT-3	Section 5.2
----------------	--	----------------

**2 x YT:** 16 g bacto-tryptone, 10 g bacto-yeast extract, 5 g NaCl made up to 1 litre with deionised water and pH adjusted to 7.0 with NaOH.

Usually used for antibiotic expression after transformation of plasmid DNA.

**ψ broth:** 2% tryptone (w/v), 0.5% yeast extract (w/v), 20 mM MgSO<sub>4</sub>, 10 mM NaCl, 5 mM KCl, adjusted to pH 7.5 with KOH.

**MacConkey Galactose Agar:** The MacConkey agar powder was supplied by Difco. A solution of 4 g in 95 ml deionised water was boiled and allowed to cool to 40 °C before adding 1% D-galactose (w/v).

### 3.7 Bacterial Cultures

Liquid cultures were grown at 37 °C with vigorous shaking, either to log phase of growth or overnight (as described in DNA preps).

Bacteria were grown on agar plates at 37 °C overnight, with the plates inverted.

Plasmid cultures to be stored were made into glycerol stock solutions, mixing 1 ml of the culture with 1 ml 2% peptone, 40% glycerol and kept at -70 °C.

Antibiotics were added to both liquid cultures and plates, as required for selection (see Table 3.5).

### 3.8 Preparation of competent *E.coli* cells

Competent cells for the transformation of plasmid DNA were prepared using the following methods:

#### 3.8.1 Calcium chloride method (CaCl<sub>2</sub>)

10 ml of L-broth was inoculated with the strain of interest and shaken overnight at 37 °C. A 19 ml solution of prewarmed L-broth was inoculated with 1 ml of this starter culture and grown with vigorous shaking until OD<sub>600</sub> of 0.4-0.6 was reached, then spun to 10,000 rpm in a pre-chilled Beckmann JA-20 rotor at 4 °C. Once the speed was reached the brake was applied immediately. The supernatant was removed and the pellet was gently suspended in 10 ml ice cold 50 mM CaCl<sub>2</sub>. The solution was respun as before and the supernatant removed. The cells were resuspended in 10 ml ice cold 50 mM CaCl<sub>2</sub> and stored at 4 °C, where they

remained competent for 48 hours. Alternatively, aliquots were frozen in liquid nitrogen for long term storage at -70 °C.

### **3.8.2 Rubidium chloride method (RbCl)**

**TFB I:** 30 mM KOAc, 100 mM RbCl, 10 mM CaCl<sub>2</sub>·2H<sub>2</sub>O, 50 mM MnCl<sub>2</sub>·4H<sub>2</sub>O, 15% (v/v) glycerol, pH 5.8 with AcOH.

**TFB II:** 10 mM MOPS, 75 mM CaCl<sub>2</sub>·2H<sub>2</sub>O, 10 mM RbCl, 15% (v/v) glycerol, pH 6.8 with HCl or KOH.

20 ml of  $\psi$  broth was inoculated with the desired strain. Using 4 ml of this culture, a 200 ml pre-warmed solution of  $\psi$  broth was inoculated and grown with shaking at 37 °C until an OD<sub>600</sub> of 0.4-0.6 (approximately 1-2 hours). The cells were then chilled on ice before spinning in a J2-21 rotor at 5000 rpm for 10 minutes at 4 °C. The supernatant was decanted and the pellet was resuspended in 10 ml of ice cold TFB I and kept on ice for 30 minutes. Following another 5000 rpm spin for 10 minutes at 4 °C the supernatant was discarded and the pellet was resuspended in 8 ml ice cold TFB II, and kept on ice for a 15 minute period. 200  $\mu$ l aliquots were removed, dispensed into Nunc tubes and frozen in liquid nitrogen for storage at -70 °C.

### **3.9 Transformation of *E. coli* with plasmid DNA**

0.1  $\mu$ g of plasmid DNA was added to 100  $\mu$ l of competent *E. coli* cells and gently mixed with a yellow pipette tip. The cell suspension was heat shocked at 37 °C for 5 minutes then placed in ice for 15 minutes. Five volumes of 2  $\times$  YT broth was added to the solution which was then incubated at 37 °C for 2 hours to allow for antibiotic expression. A 100  $\mu$ l aliquot of the culture was spread on agar plates containing appropriate antibiotics for selection.

### **3.10 Phenol Extraction**

The organic extraction of proteins from protein/DNA mixtures was carried out by adding 1 volume of phenol (containing 0.1% 8-hydroxyquinoline and equilibrated



with 0.5 M Tris-HCl pH 8). The sample was thoroughly mixed by vortexing, and spun at 14,000 rpm in a bench top centrifuge for 3 minutes. The DNA-containing aqueous layer (the upper layer) was recovered and re-extracted with more phenol if required. A further extraction of the aqueous layer with chloroform ensured removal of any residual phenol from the sample.

### **3.11 Ethanol Precipitation of DNA**

**1 x TE buffer:** 10 mM Tris-HCl pH 8.2, 1 mM EDTA.

The sodium ion concentration of the DNA solution was adjusted to 0.30 M with 3 M sodium acetate pH 5.2. To this solution was added 2 volumes of ethanol and after thorough mixing the solution was incubated at -20 °C for 30 minutes. The DNA was pelleted by spinning in a Eppendorf microfuge tube at 14,000 rpm for 30 minutes at 4 °C in bench top centrifuge. The pellet was washed with 70% ethanol and re-spun for 5 minutes before drying at room temperature for 5 minutes, then re-suspending in TE buffer.

### **3.12 Small scale preparation of plasmid DNA**

The following procedure is adapted from Birnboim and Doly (1979) and utilises alkaline lysis to disrupt the cells.

**Doly I:** 50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA.

**Doly II:** 200 mM NaOH, 1% SDS (w/v) made fresh.

**Doly III:** 0.6 vol. 5 M KOAc, 0.115 vol. AcOH, 0.285 vol. H<sub>2</sub>O.

#### **3.12.1 Alkaline lysis method**

1.5 ml cultures were grown for 12 hours at 37 °C until they reached stationary phase. The cells were pelleted by spinning at 14,000 rpm in a bench top centrifuge for 2 minutes, then re-suspended in 200 µl Doly I containing 4 mg/ml lysozyme and left at room temperature for 5 minutes. To this solution was added 400 µl Doly II. After mixing with inversion the mixture was incubated on ice for 5 minutes. Then 300 µl of 7.5 M ammonium acetate was added to the solution which was mixed thoroughly before placing on ice for 10 minutes. The cell debris

was pelleted by centrifugation at 14,000 rpm for 3 minutes. The supernatant was then removed and extracted with  $3 \times 0.5$  volumes of phenol (the first extraction was gently mixed) then 0.5 volumes of 24:1 (v/v) chloroform:isoamyl alcohol. The aqueous layer was removed and the DNA was precipitated by addition of 0.75 volumes of isopropanol, and leaving at room temperature for 10 minutes. The solution was spun at 14,000 rpm for 30 minutes at 4 °C in a bench top centrifuge. The supernatant was decanted, and the pellet was washed with 70% ethanol, dried and resuspended in 50 µl of TE buffer.

### **3.12.2 'Boiling Prep'**

**STET buffer:** 8% sucrose, 5% Triton X-100, 50 mM EDTA, 50 mM Tris-HCl (pH 8.2).

1.5 ml of overnight culture was spun at 12,000 rpm in a microfuge for 30 seconds. The supernatant was discarded, and the pellet was re-dissolved in 350 µl STET buffer. To this solution was added 25 µl of freshly made 10 mg/ml lysozyme in STET. After briefly vortexing, the solution was incubated at room temperature for 5 minutes. The solution was vortexed briefly and immediately placed in boiling water for 40 seconds. Following a 15 minute spin at 4 °C the supernatant was retained. To this was added 40 µl of 3 M NaOAc and 400 µl of ice-cold isopropanol. After thorough mixing the solution was spun at 4 °C for 7 minutes. The pellet was washed with 70% ethanol, dried and re-dissolved in 50 µl  $1 \times$  TE buffer.

### **3.12.3 Qiagen Miniprep**

This method was carried out according to the manufacturers' instructions.

## **3.13 Large scale preparation of plasmid DNA**

200 ml overnight cultures were grown for 12 hours at 37 °C until reaching stationary phase. These were then spun in a JA-14 rotor at 8,000 rpm for 5 minutes at 4 °C. The supernatant was removed, and the pellet was resuspended in

4 ml Doly I and mixed with a cut-tipped pipette. The mixture was placed on ice for 5 minutes. To the mixture was added 8 ml of freshly made Doly II, and after mixing by inversion, placed on ice for 4 minutes. 6 ml of ice-cold Doly III was added, and after mixing by vortexing, the solution was kept over ice for 5 minutes before spinning in a large centrifuge using a JA-14 rotor at 18,000 rpm for 30 minutes at 4 °C. The supernatant containing the DNA was then mixed with 12 ml isopropanol and allowed to stand at room temperature for 15 minutes before spinning at 18,000 rpm for 30 minutes at 20 °C. The pellet was washed with 70% ethanol and dried at 37 °C for 30 minutes. The pellet was resuspended in 2 ml TE buffer and placed on a horizontal shaker for 2 hours at 37 °C. A solution of 5 g of caesium chloride was added to 3 ml of water and placed in a Beckmann ultracentrifuge tube containing 270 µl of a 15 mg/ml solution of ethidium bromide and the dissolved DNA solution. The tubes were filled with liquid paraffin and heat-sealed before being weighed to balance within 0.1 g, and placed in a Beckman Ti70 fixed-angle rotor. The samples were spun in an ultracentrifuge at 49,000 rpm for 16 hours at 15 °C. The DNA was visualised using a long wave (365 nm) UV transilluminator. Usually two bands were present, the upper band consisting of chromosomal and nicked plasmid DNA, the lower band being the supercoiled plasmid DNA.

Using a 1 ml syringe fitted with a 21 gauge needle, approximately 0.5 ml of supercoiled plasmid solution was removed and placed in a test tube. Excess ethidium bromide in the sample was removed using three extractions with 0.8 volumes of *n*-butanol. The DNA solution was then placed in a Corex™ tube along with 3 volumes of deionised water and 8 volumes of ethanol, and after thorough mixing was incubated at 4 °C for 30 minutes. The solution was then spun in a JA-20 rotor at 7,000 rpm for 30 minutes at 4 °C. The pellet was then washed with 70% ethanol, dried and resuspended in 500 µl of TE buffer. The DNA was stored at -20 °C.

### 3.14 UV spectroscopy

The concentration of plasmid DNA solutions and of oligonucleotides was determined by measuring  $A_{260}$  with a UV/visible spectrophotometer (Shimadzu). The following assumption was made when measuring the DNA concentration, that an absorbance of 1 at 260 nm corresponds to a DNA concentration of 50  $\mu\text{g/ml}$  for double-stranded DNA.

For oligonucleotides, the absorbance at 260 nm was used to estimate the DNA concentration using the Beer-Lambert Law

$$A = \epsilon \cdot c \cdot l$$

Where  $A$  = absorbance at 260 nm

$\epsilon$  = extinction co-efficient, estimated at 10,000  $\text{l mol}^{-1} \text{cm}^{-1}$  per nucleotide

$c$  = molar concentration

$l$  = path length of cell (usually 1 cm)

### 3.15 Restriction endonuclease digestion of DNA

Digestion was carried out in the supplier's recommended buffer, usually BRL React buffer or New England Biolabs buffer, with a 10 fold excess of enzyme to ensure complete digestion of the DNA (10 U/ $\mu\text{g}$  DNA). For cloning purposes, usually 1-5  $\mu\text{g}$  plasmid DNA was digested, and for construction of artificial supercoiled substrates, 50  $\mu\text{g}$  was cut and higher concentration enzymes were used.

In a typical digestion, the DNA was cut in 20  $\mu\text{l}$  of the correct buffer for 1 hour at 37 °C, and the reaction was stopped either by heat inactivation of the enzyme at 70 °C for 10 minutes, by the addition of SDS loading buffer, or by extracting with 1 volume of phenol.

Double digests were carried out by cutting first with one enzyme, then heating to 70 °C for 10 minutes. After adjusting the pH and/or concentration of sodium ions, if required, the second enzyme was used and the mixture was incubated for 60 minutes at 37 °C as before.

### **3.16 5'-Phosphorylation of DNA**

**1 × T4 kinase buffer:** 50 mM Tris-HCl (pH 7.5), 10 mM MgCl<sub>2</sub>, 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA.

This method allowed the introduction of either a labelled or unlabelled phosphate group at the 5' end of dephosphorylated linear DNA fragments or synthetic oligonucleotides. The phosphate group is derived from either unlabelled ATP or [ $\gamma$ -<sup>32</sup>P]ATP.

Typically 2 pmol of DNA was added to 1 × T4 kinase buffer (20  $\mu$ l) with either 10  $\mu$ Ci [ $\gamma$ -<sup>32</sup>P]ATP (8000  $\mu$ Ci/mmol) or 1 mM ATP, followed by 1 unit of T4 kinase. After incubation at 37 °C for 45 minutes, the reaction was stopped by heating at 70 °C for 10 minutes. The phosphorylated DNA fragment was recovered by ethanol precipitation.

### **3.17 Filling in 3' recessed ends of DNA fragments**

Following restriction endonuclease digestion, it is sometimes desirable to fill in the 3'-recessed end, for a blunt-ended ligation or to remove a restriction enzyme site from plasmid DNA. This was achieved by adding the Klenow fragment from DNA polymerase I (1 unit) and the appropriate dNTPs (100  $\mu$ M final) to the digested DNA, and incubating for 30 minutes at 37 °C.

### **3.18 3'-end labelling of DNA fragments**

This was used to incorporate a radionuclide label into the DNA after a restriction digest, at the 3' recessed end created by that digest.

1-2 µg of DNA was cut with a restriction endonuclease generating a 3' recessed end. 10 µCi of an [ $\alpha$ - $^{32}$ P]dNTP (3000 µCi/mmol) and the other dNTPs at a final concentration of 10 µM were added to the digested DNA, along with 1 unit of the Klenow fragment of DNA polymerase I. The sample was incubated at room temperature for 20 minutes. Cold dNTPs were then added to a final concentration of 100 µM each and the solution was kept at room temperature for 20 minutes. The reaction was stopped by phenol and chloroform extraction and the 3'-labelled DNA was recovered by ethanol precipitation.

### 3.19 Ligation of DNA restriction fragments

**1 × BRL ligation buffer:** 50 mM Tris-HCl (pH 7.6), 10 mM MgCl<sub>2</sub>, 1 mM ATP, 1 mM DTT, 5% (w/v) polyethylene glycol-8000.

Ligations for cloning purposes used 1-2 µg (approx. 1 pmol) of DNA in a volume of 20 µl with 1 × BRL ligation buffer and 1 unit of T4 DNA ligase. Samples were incubated at room temperature overnight, then used to transform competent *E.coli* cells.

If more efficient transformation was required, it was sometimes necessary to precipitate the ligation mixture with ethanol before transforming the competent cells.

### 3.20 Annealing oligonucleotides

Single-stranded oligonucleotides were annealed in 10 µl of 100 mM NaCl, 10 mM Tris-glycine (pH 9.4). Unlabelled oligonucleotides were annealed in a 1:1 ratio, whereas when annealing a [ $\gamma$ - $^{32}$ P]ATP labelled strand to an unlabelled strand, a ratio of 2:1 labelled to unlabelled was used. This ensured complete annealing of the labelled strand. For the modified strands carrying the protected thiol group, a ratio of 1:1.5 wild-type strand to labelled modified strand was used. Typically, 2 pmol of oligonucleotide (labelled and/ or modified) was annealed to the corresponding strand in 10 µl of annealing buffer. The mixture was heated to 70 °C for 10 minutes, then allowed to cool to room temperature.

### 3.21 Agarose gel electrophoresis

**1 × TAE buffer:** 40 mM Tris-acetate (pH 8.2), 20 mM sodium acetate, 1 mM Na<sub>2</sub>EDTA.

Agarose gel electrophoresis was used to separate and visualise DNA from single colonies, restriction digests, and recombination assays. Normally 0.7%, 1.2% and 1.5% agarose gels were made by mixing BRL Ultrapure agarose in 250 ml of 1 × TAE buffer and heating for 3-4 minutes in a 950 W microwave oven until the agarose was completely dissolved. The hot solution was allowed to cool for 20 minutes before pouring.

FMC low melting point 'Seaplaque' agarose was employed when it was necessary to recover DNA bands from the gel, i.e. for cloning or for making artificial supercoiled substrates. Seaplaque agarose was left to set at 4 °C.

A 150 × 200 mm horizontal slab gel was prepared by pouring the hot agarose solution (250 ml) into a gel formed with a 22 well comb close to one edge. Around 3 litres of 1 × TAE was required in the buffer reservoir tank. The gels were usually run for 3 hours at 100 V or at 40 V overnight.

Smaller gel kits were used to quickly check restriction digests. These 'minigels' require 30 ml of agarose gel, 400 ml 1 × TAE and were run for 1 hour at 80-100 V.

### 3.22 'Single Colony Gel' Analysis

**Single colony lysis buffer:** 2% ficoll (w/v), 1% SDS (w/v), 0.05% orange G (w/v), 0.01% bromophenol blue (w/v) in 1 x TAE buffer.

This technique was used to confirm that transformations with ligated DNA had been successful. Single bacterial colonies of interest were made into 1 cm<sup>2</sup> patches with a sterile toothpick on a fresh selective agar plate. The plates were grown overnight. A sample of the patch was picked up using a sterile toothpick,

mixed in 1 × single colony lysis buffer, and incubated at room temperature for 20 minutes. The samples were spun in a bench top centrifuge at 14,000 rpm for 20 minutes. The supernatant was directly loaded onto a 1.5% agarose gel.

### **3.23 Purification of DNA from low melting point gels**

Low melting point agarose 'Seaplaque GTG' was used to purify large linear DNA fragments, usually following restriction digestion. After visualising the bands of interest using the long wave (365 nm) transilluminator, they were excised from the gel using a clean scalpel and transferred to an eppendorf tube.

The gel chip was melted at 65-70 °C for 10 minutes, then allowed to cool briefly before extracting with 1 volume of phenol. At this point the solution was mixed gently by inversion to minimise emulsification. The solution was spun in a microfuge at 14,000 rpm for 5 minutes, then the aqueous layer was removed and placed in a fresh eppendorf tube. The solution was extracted further with another 2 × 1 volume of phenol, mixing by vortexing before spinning, then with 1 volume of chloroform. The DNA was ethanol-precipitated. The pellet was washed with 70% ethanol and dissolved in 1 × TE buffer.

### **3.24 Polyacrylamide Gel Electrophoresis**

Different gel compositions and running conditions were used to analyse proteins, fragments of DNA, oligonucleotides and protein/DNA crosslinking.

**1 × TBE:** 89 mM Tris base, 89 mM boric acid, 0.2 mM EDTA.

**1 × Tris-glycine:** 50 mM Tris-glycine (pH 9.4), 0.1 mM EDTA. The buffer was made as a 10 × stock solution of 500 mM Tris, 100 mM glycine, 10 mM EDTA. The pH of this was approximately 9.4 and was not adjusted further.

#### **3.24.1 SDS-polyacrylamide gels**

**1 × Laemmli running buffer:** 25 mM Tris base, 20 mM glycine, and 0.1% SDS (w/v).



These were used to check the induction of resolvase protein prior to purification, by looking at whole cell samples. They were also used to monitor the purification and yield of resolvase mutants. A Biorad Mini-Protean II dual slab gel kit was used, usually to run two gels simultaneously. A discontinuous buffer system was used which involved running the samples through gels of different percentage acrylamide and different ionic strength.

A resolving gel comprising 15% polyacrylamide, 0.375 M Tris-HCl (pH 8.8), 0.1% SDS was made by mixing 5 ml 30% acrylamide:0.8% bisacrylamide (w/v), 2.5 ml 1.5 M Tris-HCl (pH 8.8), 100  $\mu$ l 10% SDS, 2.35 ml dH<sub>2</sub>O, 50  $\mu$ l 10% APS and 5  $\mu$ l TEMED. The resolving gel was poured, overlain with isopropanol, and allowed to polymerise until set. The overlay solution was poured off, and the top of the gel was washed with de-ionised water and dried. The stacking gel, comprising 4% polyacrylamide, 0.125 M Tris-HCl (pH 6.8), 0.1% SDS was made by mixing 1.3 ml 30: 0.8 acrylamide:bisacrylamide (w/v), 2.5 ml 0.5 M Tris-HCl (pH 6.8), 100  $\mu$ l 10% SDS, 6.1 ml dH<sub>2</sub>O, 50  $\mu$ l 10% APS and 10  $\mu$ l TEMED. This solution was poured onto the resolving gel that had already set and a 10 well comb was inserted.

During the protein purification steps, aliquots of 1/200th of the total volume of the resolvase-containing fractions were taken for analysis. Before running the samples, 1-2  $\mu$ l of  $\beta$ -mercaptoethanol was added to the loading buffer (Section 3.28) to ensure the protein would be fully reduced. Gels were run at 200 V for 45 minutes.

#### **3.24.2 Non-denaturing polyacrylamide gels**

This technique was used to separate and isolate small fragments of DNA (< 300 bp) and to analyse oligonucleotides.

Adjustable slab gel kits from CBS Scientific Co. were used. 160 mm x 180 mm glass plates were clamped with a rubber seal to prevent any leaks during pouring.

The gel solution was poured between the plates, then a 15-well comb was inserted at the top of the gel and held in place with a clamp until set. After the gel had polymerised, it was clamped to the electrophoresis kit, and after the running buffer reservoirs were filled the gel comb was removed. Prior to loading the samples, the wells were washed thoroughly with running buffer. Typically gels were run at 200 V at room temperature for 3-4 hours. The DNA was visualised using ethidium staining, autoradiography or phosphor-imaging.

Gels of different percentages of acrylamide were made by changing the water to acrylamide ratio. Typically 40 ml 10% (w/v) gel was made from 14 ml 30% 30:0.8 acrylamide: bisacrylamide (w/v), 4 ml 10 × TBE (or Tris-glycine), 22 ml dH<sub>2</sub>O, 360 µl 10% ammonium persulphate (APS) and 18 µl TEMED.

### **3.24.3 Denaturing polyacrylamide gel electrophoresis**

For the analysis of crosslinking reactions the recipe for a 10% gel described in the previous Section was used. A 10% Tris-glycine gel was prepared, with SDS added to both the gel and the running buffer to a final concentration of 0.1%.

For the purification of synthetic oligonucleotides, gels containing urea were prepared. A 75 ml, 15% gel was made by mixing 31.5 g urea, 7.5 ml 10 × TBE, 22.5 ml 40% 19:1 acrylamide: bisacrylamide (w/v), 21.5 ml dH<sub>2</sub>O. The mixture was warmed until the urea dissolved, then 480 µl 10% ammonium persulphate (APS) and 27 µl TEMED were added. The gel was poured between two plates (320 mm × 180 mm) and left to polymerise for 1 hour. The plates were loaded into the CB8 polyacrylamide gel kit. The wells were washed thoroughly and the gel was preheated at 400-500 V (15 W) prior to sample loading.

To the sample was added 1 volume of formamide loading buffer (Section 3.28), then the sample was heated to 70 °C for 5 minutes and loaded on to the gel and run at 500 V (constant voltage) for 2-3 hours. The gel was removed and stained with 70 ml dH<sub>2</sub>O: 20 ml isopropanol:10 ml 0.1% “Stains-all” (Section 3.29.4) in

formamide for 10 minutes or until the bands could be seen. After washing for 5 minutes with deionised water, the bands were removed using a scalpel and placed in a Nunc tube. After adding 1 ml 1 × TE buffer and crushing with a glass rod, the samples were placed on a rotary mixer at 37 °C overnight. The mixture was passed through a Co-star microspin 0.45 µm filter and the eluate was dried directly in a gyrovap, or ethanol precipitated before re-dissolving in 1 × TE buffer.

#### **3.24.4 DNA sequencing gels**

These were used to analyse the dinitrophenyl-modified (DNP) oligonucleotides (Section 3.38). A 100 ml 10% acrylamide-7 M urea gel was made by mixing 40 g urea, 10 ml 10 × TBE, 25 ml 40% 19:1 acrylamide: bisacrylamide (w/v), 20 ml dH<sub>2</sub>O and 40 µl TEMED in a glass beaker. A Gibco BRL sequencing gel kit was used. The solution was warmed until the urea dissolved, then 800 µl 10% APS was added. The gel solution was poured between 410 mm × 330 mm plates and allowed to polymerise. The gel was pre-run at 80 W until the temperature reached 48 °C, before sample loading. The samples were heated to 70 °C in 1 volume of formamide loading buffer (Section 3.28) before loading. The gel was run at 60 W for 1 hour.

#### **3.25 DNA synthesis**

The reagents used were supplied by Cruachem. Oligonucleotides were synthesised using an Applied Biosystems Model 392 DNA synthesiser. The synthesis of the oligonucleotides is based on standard cyanoethyl phosphoroamidite chemistry (reviewed in Goodchild, 1990). For the synthesis of the phosphorothioate-containing oligonucleotides, an additional step was used to replace the O with S. Prior to the oxidation step, the column was removed from the machine and washed with 1 ml of the sulfurization reagent 3*H*-1,2-benzodithiol-3-one, 1,1-dioxide (10 mM) in acetonitrile for 2 minutes.

### 3.26 DNA molecular weight standards

These were used on agarose and polyacrylamide gels, usually the Gibco BRL 1 kb ladder.

### 3.27 Purification of DNA from polyacrylamide gels

The band of interest was excised from the gel using a clean scalpel. The gel chip was placed in a Nunc tube, crushed with a glass rod and placed in 1 ml of  $1 \times$  TE buffer. The sample was placed on a rotary mixer overnight at 37 °C. The excess acrylamide was removed by spinning the solution through a Co-star spinX 0.22  $\mu$ m column. The DNA solution was either precipitated with ethanol or concentrated in a gyrovap and re-dissolved in TE buffer.

### 3.28 Loading buffers

**Formamide loading buffer:** 80% deionised formamide (v/v), 10 mM EDTA (pH 8.0), 1 mg/ml xylene cyanol FF, 1 mg/ml bromophenol blue.

One volume of the above was added to the oligonucleotide solutions prior to running on a denaturing polyacrylamide gel. To denature the oligonucleotides prior to loading on a gel, the samples were heated in the formamide buffer for 10 minutes at 55 °C.

**SDS loading buffer:** 50% glycerol, 10 mM Tris-HCl (pH 8.2), 1% SDS, 0.01% bromophenol blue in  $1 \times$  TAE buffer.

0.25 volumes of the above was added to samples prior to running on agarose or polyacrylamide gels. To remove protein from crosslinked DNA, proteinase K at 10 mg/ml was added to this buffer.

**Laemmli loading buffer:** 50 mM Tris-HCl (pH 6.8), 1% SDS (w/v), 10% glycerol (v/v), 0.1% bromophenol blue.

An aliquot from a protein purification step or 5-10  $\mu$ l of a purified resolvase dilution was added to 20  $\mu$ l of Laemmli loading buffer. 1-2  $\mu$ l of 100%  $\beta$ -mercaptoethanol was added to the samples, which after thorough vortexing were boiled for 10 minutes prior to loading on the gel.

### **3.29 Gel staining and visualisation**

#### **3.29.1 Ethidium staining of agarose and polyacrylamide gels**

Unlabelled DNA was visualised using ethidium bromide at a final concentration of 0.6 µg/ml in 1 litre of running buffer. The tray was shaken gently on a R100 Luckham Shaker for 30 minutes, then the gel was rinsed in de-ionised water and allowed to de-stain for 30 minutes to remove excess ethidium bromide. For general DNA analysis and resolution reactions, the stained DNA was visualised using short-wave 254 nm UV light on a ultraviolet transilluminator (UVP). Bands that were required for further experiments were visualised on a long-wave 365 nm transilluminator to reduce photochemical damage to the DNA.

#### **3.29.2 Visualisation of radiolabelled DNA**

Gels were placed on 3 layers of Whatman 3MM paper and were dried at 80 °C for 1-2 hours under vacuum in a Biorad slab gel drier. The dried gels were either exposed against a Fuji RX100 film then developed using an X-OMAT automated processor (Kodak), or against a Fuji phosphoimaging plate, then scanned on a Fuji BAS-1500 phosphor-imaging system.

#### **3.29.3 Visualisation of proteins**

**Coomassie Stain:** 1% Coomassie Blue R250 (w/v), 50% MeOH (v/v), 10% AcOH (v/v).

**Coomassie Destain:** 10% MeOH (v/v), 10% AcOH (v/v).

After electrophoresis, the gel was covered with stain solution and stained for 30 minutes with gentle agitation, then destained for at least 1 hour. The protein bands were visualised on a light box.

#### **3.29.4 Visualisation of oligonucleotides**

**“Stains-all”:** 1-ethyl-2-[3-(1-ethylnaptho[1,2-d]thiazolin-2-ylidene)-2-methylpropenyl]-naptho(1,2-d)thiazolium bromide.

The gel was placed in a shallow glass tray containing 70 ml of water and 20 ml of isopropanol. To this was added 10 ml of a 0.1% (w/v) solution of 'Stains-all' in formamide, and after gentle agitation for 5-10 minutes the bands appeared. After destaining by rinsing in water for 5 minutes, the bands were excised, crushed with a glass rod, then soaked in 1 ml of  $1 \times$  TE and placed on a rotary mixer overnight. Generally, the slowest moving strong band on the gel corresponded to the full length oligonucleotide.

### **3.30 Tn3 resolvase**

Wild-type resolvase R17f33 was a gift from M. Boocock; its concentration was 48  $\mu$ M (Blake, 1993; Watson, 1994). Wild-type Tn3 resolvase and the cysteine mutants (Section 3.30 and 4.10) were made into solutions of different concentrations by diluting in resolvase dilution buffer. These dilutions were annotated  $2^{-1}$ ,  $2^{-2}$ ,  $2^{-3}$ , etc indicating a two-fold, four-fold and eight-fold dilution respectively.

**Resolvase dilution buffer:** 10 mM Tris-HCl (pH 7.5), 1 M NaCl, 50% glycerol (v/v), 0.1 mM EDTA, 0.1 mM DTT (1 mM DTT for cysteine mutants).

### **3.31 *In vitro* recombination**

The reactions were carried out in the following buffers:

**C8.2:** 50 mM Tris-HCl (pH 8.2), 10 mM  $\text{MgCl}_2$ , 0.1 mM EDTA.

**C9.4:** 50 mM Tris-HCl (pH 9.4), 10 mM  $\text{MgCl}_2$ , 0.1 mM EDTA.

**MD10:** 50 mM glycine-NaOH (pH 10), 5 mM spermidine. $3\text{HCl}$ , 5 mM EDTA.

For resolution reactions, 0.05 volumes of diluted resolvase was added to 0.8  $\mu$ g of DNA in 40  $\mu$ l of reaction buffer. For the cysteine-modified proteins 0.1 volumes of protein were used. Reactions were incubated at 37 °C for 60 minutes, then the protein was heat-inactivated at 70 °C for 10 minutes. The sample was divided into  $2 \times 20$   $\mu$ l aliquots, one of which was analysed by restriction digestion. 0.25 volumes of SDS-loading buffer was added to the samples before loading on an

agarose gel. 1 mM DTT was added to all the buffers used for reactions of cysteine mutants.

### 3.32 Band shift analysis

**Binding buffer:** 10 mM Tris-glycine (pH 9.4), 0.1 mM EDTA, 1 mM MgCl<sub>2</sub>, 10% glycerol (v/v).

Non-specific carrier DNA, usually pUC71K (10-25 µg/ml), was mixed with end-labelled *res* DNA or site I oligonucleotides (400-600 cps) in 200-300 µl binding buffer. The reactions were divided into 20 µl aliquots and 0.05 volume of wild-type protein and cysteine mutants was added. The samples were incubated at 37 °C for 15 minutes, then loaded onto a 6% non-denaturing polyacrylamide gel (Section 3.24.2) and run at 200 V for 4 hours.

### 3.33 Preparation of competent BL21(DE3)pLysS cells

A glycerol stock solution of BL21(DE3)pLysS was available in the lab (Arnold, 1999). A toothpick was used to remove a small sample of these cells which were streaked on agar containing 25 µg/ml chloramphenicol and grown overnight at 37 °C. A single colony was removed using a sterile toothpick and placed in a tube containing 30 ml L-broth. This sample was grown, with shaking, overnight at 37 °C.

This overnight culture was used to prepare competent cells of BL21(DE3)pLysS using the CaCl<sub>2</sub> method (Section 3.8.1). These competent cells were left on ice overnight, then transformed with the pMS95-99 derivatives, which encode the resolvase C-terminal cysteine mutants. The transformations were plated on agar containing 50 µg/ml ampicillin and 25 µg/ml chloramphenicol and grown overnight at 37 °C. Colonies were streaked onto agar, with the same antibiotic selection as above, and grown overnight at 37 °C.

### **3.34 Expression of the mutants in the expression strain BL21(DE3)pLysS**

**Cell re-suspension buffer:** 20mM Tris-HCl pH 7.5, 10mM MgCl<sub>2</sub>, 0.1mM EDTA.

The transformant clones were streaked to single colonies, and these were used to inoculate an overnight starter culture. 1 ml of this starter culture was used to inoculate a 400 ml shake culture in L-broth with 50 µg/ml kanamycin. The cells were grown at 37 °C until the A<sub>600</sub> was between 0.45-0.60. The cells were then immediately induced with 0.5 mM IPTG and shaken vigorously at 37 °C for 3.5 hours. The cells were harvested into pre-weighed 200 ml centrifuge tubes and centrifuged at 9 K for 10 minutes at 4 °C. The supernatant was removed and the spin step was repeated, to ensure that all the supernatant was discarded. The cells were washed in 100 ml of ice-cold re-suspension buffer on ice using a glass rod and gently pipetting up and down. The cells were spun at 9 K for 10 minutes as before and the supernatant was thoroughly removed. The dry pellets were weighed and stored at -70 °C.

### **3.35 Purification of cysteine mutants of Tn3 resolvase**

The purification is based on a denaturing method (Boocock and Weinwieser, pers. comm.). All steps were performed at 4 °C unless otherwise stated, and after each wash the solutions were spun at 18 K for 15 minutes at 4 °C in a JA-20 Beckmann rotor. At each step of the protocol, samples were taken to be run on Laemmli gels (Section 3.24.1). The quantities in the following correspond to starting with approximately 1-2 g dry cells.

**TMDEP:** 20 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 5 mM DTT, 0.1 mM EDTA, 1.2 mM PMSF, 1% EtOH.

**TDEP:** 20 mM Tris-HCl pH 7.5, 5 mM DTT, 0.1 mM EDTA, 1.2 mM PMSF, 1% EtOH.

**Buffer A:** 25 mM Tris-HCl pH 7.5, 25 mM NaCl, 0.1 mM EDTA, 5 mM DTT, 6M urea.



**Buffer B:** 25 mM Tris-HCl pH 7.5, 1.0 M NaCl, 0.1 mM EDTA, 5 mM DTT, 6 M urea.

The cells were thawed over ice for 20 minutes, then placed at room temperature for 5 minutes before being re-dissolved in 15 ml TMDEP. Initially only 0.5 mM PMSF was added to the initial re-suspension buffer, with a further 0.5 mM added after the first sonication. The mixture was sonicated using a Vibra-cell VC100 for  $3 \times 15$  seconds, then  $2 \times 30$  seconds, at 40% amplitude using a button probe.

Between sonication steps, 10 minutes cooling with a salt-ice mixture ensured that the temperature was maintained below 12 °C. The mixture was centrifuged at 18 K for 15 minutes at 4 °C to precipitate the insoluble cell debris concomitantly with the resolvase.

The pellet was re-suspended in 10 ml TMDEP containing 100 mM NaCl, using a glass rod and vortexing. The solution was then transferred to a Dounce homogenizer, where it was thoroughly homogenized around 100 times. The solution was spun as before. At this stage the pellet should contain resolvase protein. The pellet was re-suspended in 10 ml TMDEP containing 2M NaCl using a glass rod, and vortexing. Homogenisation was as in the low salt wash step, although this time the homogenisation was more vigorous at 5-10 minute intervals over a period of 30 minutes. The solution was then spun as before. The supernatant was retained as it may contain some soluble resolvase at this stage.

The pellet was then re-suspended in 10 ml TDEP containing 6 M urea. As before, the pellet was initially broken up using a glass rod and vortexing, before being transferred to the Dounce homogeniser. This time, it was homogenised over a period of 90 minutes, allowing the solution to warm up to around 5-10 °C to encourage solubilisation. Following the spin, most of the resolvase was contained in the supernatant which was retained. As a final wash the pellet was re-suspended in 10 ml TDEP containing 6 M urea and 1% SDS at room temperature for 10 minutes. Following the spin both the pellet and supernatant were retained.

The fraction that contained the most resolvase, usually the 6 M urea fraction, was dialysed against 1 litre of TDEP/1 M NaCl for 6 hours, to allow for refolding/renaturation of the protein. This dialysate was then immediately dialysed against 1 litre of TMDEP to allow the protein to precipitate.

The precipitated protein was then centrifuged at 18 K for 15 minutes at 4 °C and the pellet was re-suspended in 10 ml buffer A. The protein was purified on a Millipore Waters 650E Advanced Protein Purification System using cation exchange resin SP sepharose in a 6 ml Pharmacia column. The column was washed thoroughly with 2 M NaCl, 25 mM Tris-HCl pH 7.5 before being re-equilibrated using buffer A. The protein solution was freshly spun at 18 K for 15 minutes at 4 °C before being loaded manually onto the column using a peristaltic pump (Pharmacia LKB Pump-P1). The column was washed with several volumes of buffer A before the salt gradient was applied. The column was washed with a linear gradient of 25-250 mM NaCl for 25 minutes, then 250-1000 mM NaCl for a further 25 minute period.

The fractions collected were sampled on a Laemmli gel, and the protein-containing fractions were dialysed against 1 litre TMDEP containing 1 M NaCl for 6 hours followed by dialysis against low salt TMDEP as before.

The second purification step was using a Pharmacia MonoS™ cation exchange column. The protein from the low salt dialysis was re-suspended in 2 ml buffer A and applied to the column using the injector loop. This time, a linear gradient of 25-1000 mM NaCl was run over 1 hour to achieve better separation of the proteins. The fractions were sampled on Laemmli gels as before, and the purest fractions were dialysed to refold and precipitate the resolvase.

The final dialysate was spun down for 5 minutes in the cold room, ensuring that all traces of supernatant were removed from the pellets. The pellets were re-suspended in either 200 µl or 400 µl of a solution containing 20 mM Tris-HCl pH

7.5, 0.1 mM EDTA, 2 M NaCl and 5 mM DTT for 20 minutes over ice, pipetting up and down and vortexing to encourage solubilisation. The solutions were spun in a bench top microfuge at 4 °C and the supernatants were transferred to fresh Nunc tubes. After the addition of an equal volume of glycerol to the final protein solution, the samples were mixed thoroughly, spun briefly and stored at –20 °C.

### 3.36 Construction of artificial supercoiled substrates

**Shore ligase buffer:** 10 mM Tris-HCl (pH7.5), 50 mM NaCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT and 1 mM ATP.

Artificial supercoiled substrates were made by ligating modified site I double-stranded oligonucleotides into the *EcoRI* and *SstI* sites of pMSH2 (Section 6.2). The substrate pMSH2 was digested with *EcoRI* and *SstI*, and the large linear fragment was purified by gel electrophoresis using low melting point agarose (Section 3.21). This linear fragment was ligated to a double-stranded site I oligonucleotide labelled at one 5' end with  $\gamma$ -[<sup>32</sup>P]ATP and T4 kinase. The 5' end of the other strand was phosphorylated with unlabelled ATP and T4 kinase (Section 3.16).

The procedure involves two ligation steps. The first ligation was in a volume of 100 µl and contained ligation buffer (Shore *et al.*, 1981) and 150 U/ml T4 DNA ligase (BRL); the DNA concentration was about 300 µg/ml. After incubating at room temperature for 15 hours, the ligase was inactivated by heating at 70 °C for 10 minutes. The mixture, which was now concatameric linear DNA, was digested with 800 U/ml *SstI* by incubating at 37 °C for 1 hour, and then the *SstI* was inactivated by heating at 70 °C for 10 minutes. The mixture was diluted with 1 × ligation buffer to a final DNA concentration of 3 µg/ml, and then 5.1 µg/ml ethidium bromide, 1 mM ATP and 7.5 U/ml T4 DNA ligase was added. After ligating at room temperature for 15 hours, the solution was extracted with phenol, and DNA was ethanol-precipitated. The supercoiled, semi-synthetic plasmid was purified by electrophoresis using a 0.8% low melting point Seaplaque gel

(Section 3.21). Natural pMSH2 was run alongside the ligation products to indicate the position of the supercoiled form of the plasmid.

### **3.37 *In vivo* testing of mutant resolvase plasmids**

A *galK* assay, developed previously (Blake, 1993), was used to test the ability of the cysteine mutants of resolvase to carry out recombination *in vivo*.

Recombination *in vivo* is observed using MacConkey agar and a specifically designed test plasmid. Here pDB34 was used, which contains two full *res* sites in direct repeat, the sites flanking the galactose gene, *galK*.

The *galK* enzyme brings about phosphorylation of galactose prior to it being fermented to lactic and acetic acids. The formation of these acidic products turns the MacConkey agar plates (containing Neutral Red indicator) red. In the absence of *galK* the cells must utilise the amino acids present in the media, making ammonia as the chief by-product, raising the pH and thus turning the indicator white.

In the presence of a recombinationally-active resolvase, the test plasmid will be split into two smaller circles. The one containing the origin of replication will survive and the other, which contains the *galK* gene, will be lost. These cells will appear as white colonies on the selective media. Expression from plasmid containing a mutant resolvase that is inactive in recombination will give red colonies, due to the expression of *galK*.

CaCl<sub>2</sub> competent cells of DS941 carrying the test plasmid pDB34 were prepared using kanamycin selection for cells containing the substrate. These cells were then transformed with the mutant resolvase expression plasmids, pMS95-pMS99, which carry ampicillin resistance. After transformation, the cells were plated on MacConkey agar plates containing ampicillin and kanamycin and incubated at 37 °C overnight.

### **3.38 Synthesis of modified oligonucleotides**

The site I oligonucleotides consisted of a 40-mer top strand and a 48-mer bottom strand (sequences shown in Table 3.4 SHT1-SHB5). The DNA was synthesised using an Applied Biosystems Model 392 DNA synthesiser (Section 3.25). The synthesis was altered to allow maximum incorporation of the chemically modified base. Once the modified base was added to the column, the coupling step was increased to 300 seconds.

### **3.39 Ammonia deprotection of the modified oligonucleotides**

The oligonucleotides were cleaved from the glass support by treatment with 1 ml of 30% ammonium hydroxide for 1 hour, then stored at -70 °C to minimise degradation. Aliquots of oligonucleotides were removed and treated with an equal volume of ammonium hydroxide at room temperature for 72 hours, to ensure removal of exocyclic amides while maintaining the integrity of the DNP protecting group (Hanna and Meyer, 1996).

### **3.40 $\beta$ -mercaptoethanol deprotection of 5-thiol DNP-containing oligonucleotides**

2 pmol of the modified strand was first end-labelled with [ $\gamma$ -<sup>32</sup>P]ATP then annealed to the complementary wild-type strand in 10  $\mu$ l of annealing buffer. To this solution was added 26  $\mu$ l of 20 mM Tris-HCl (pH 8.2) and 4  $\mu$ l of fresh 100%  $\beta$ -mercaptoethanol. The reaction mixture was heated at 45 °C for a minimum of 4 hours, although an overnight deprotection (12 hours) was also carried out. The solution was spun briefly, then ethanol-precipitated by adding 10  $\mu$ g of yeast tRNA, 1/10th volume of 3 M NaOAc and 2 volumes of 100% ethanol. The mix was cooled to -20 °C for 30 minutes, then spun at 4 °C for 1 hour. The pellet was washed with 70% ethanol, centrifuged for 2 minutes at room temperature then dried in a gyrovap for 15 minutes. The pellet was redissolved in 10  $\mu$ l of annealing buffer (Section 3.20).

### 3.41 Alkylation of thiol-containing oligonucleotides

The oligonucleotides were treated with  $\beta$ -mercaptoethanol and precipitated as in Section 3.40, then dissolved in 20  $\mu$ l of TE buffer. The samples were divided into  $2 \times 10$   $\mu$ l aliquots, one of which was treated with 2  $\mu$ l of iodoacetimido-fluorescein (IAF; 10 mM in dimethyl formamide) overnight at room temperature. The solution was ethanol-precipitated as with the  $\beta$ -mercaptoethanol deprotection step (Section 3.40) with a 1 hour spin at 4 °C, and redissolved in 10  $\mu$ l TE containing 1 mM DTT. The treated and untreated samples were run on a 10% sequencing gel (Section 3.24.4).

### 3.42 Crosslinking reactions

The thiol-containing double-stranded (Section 3.20) oligonucleotides were reacted with the cysteine mutants in a binding reaction, followed by addition of crosslinking reagents:

**Binding buffer:** 10 mM Tris-glycine (pH 9.4), 0.1 mM EDTA, 1 mM  $\text{MgCl}_2$ , 10% glycerol (v/v).

Non-specific carrier DNA, usually pUC71K (10  $\mu$ g/ml) was mixed with end-labelled (double-stranded) site I oligonucleotides (360-400 cps) containing the free thiol group. A total binding mix of 360  $\mu$ l or 400  $\mu$ l was prepared for 9 or 10 40  $\mu$ l crosslinking reactions. Then 0.05 volumes of the diluted resolvase cysteine mutant were added, and the reactions were incubated at 37 °C for 15 minutes.

The reactions were then treated with oxidising agents. To the 40  $\mu$ l binding reaction, the oxidant was added, the solution vortexed, spun briefly and left at room temperature, for times ranging from 1 hour to 24 hours. A final concentration of 1 mM for glutathione and *cis*-platinum diamine dichloride (cisplatin), and 0.5 mM for sodium selenite were used. For air oxidation, the tube was left open to the air overnight.

After the oxidation step, the samples were divided into two 20  $\mu$ l aliquots, one of which was treated with 0.2 volumes of SDS-loading buffer containing proteinase K, the other with SDS-loading buffer only. After 30 minutes, the reactions were loaded onto a 10% Tris-glycine-containing 0.1% SDS (v/v) polyacrylamide gel (Section 3.24.2), and run at 200 V for 4 hours.

### **3.43 2D gel analysis of crosslinked complexes**

A binding reaction was carried out between the thiol-containing double-stranded oligonucleotides and the cysteine mutants, and the products were run out on Tris-glycine gels (Section 3.24.2). After running for 3 hours, the gel was soaked in running buffer containing 0.1% SDS for 30 minutes, transferred to a horizontal gel kit, and run for 2 hours at 100 V at right angles to the original direction.

## **Chapter 4: Synthesis and characterisation of novel cysteine mutants**

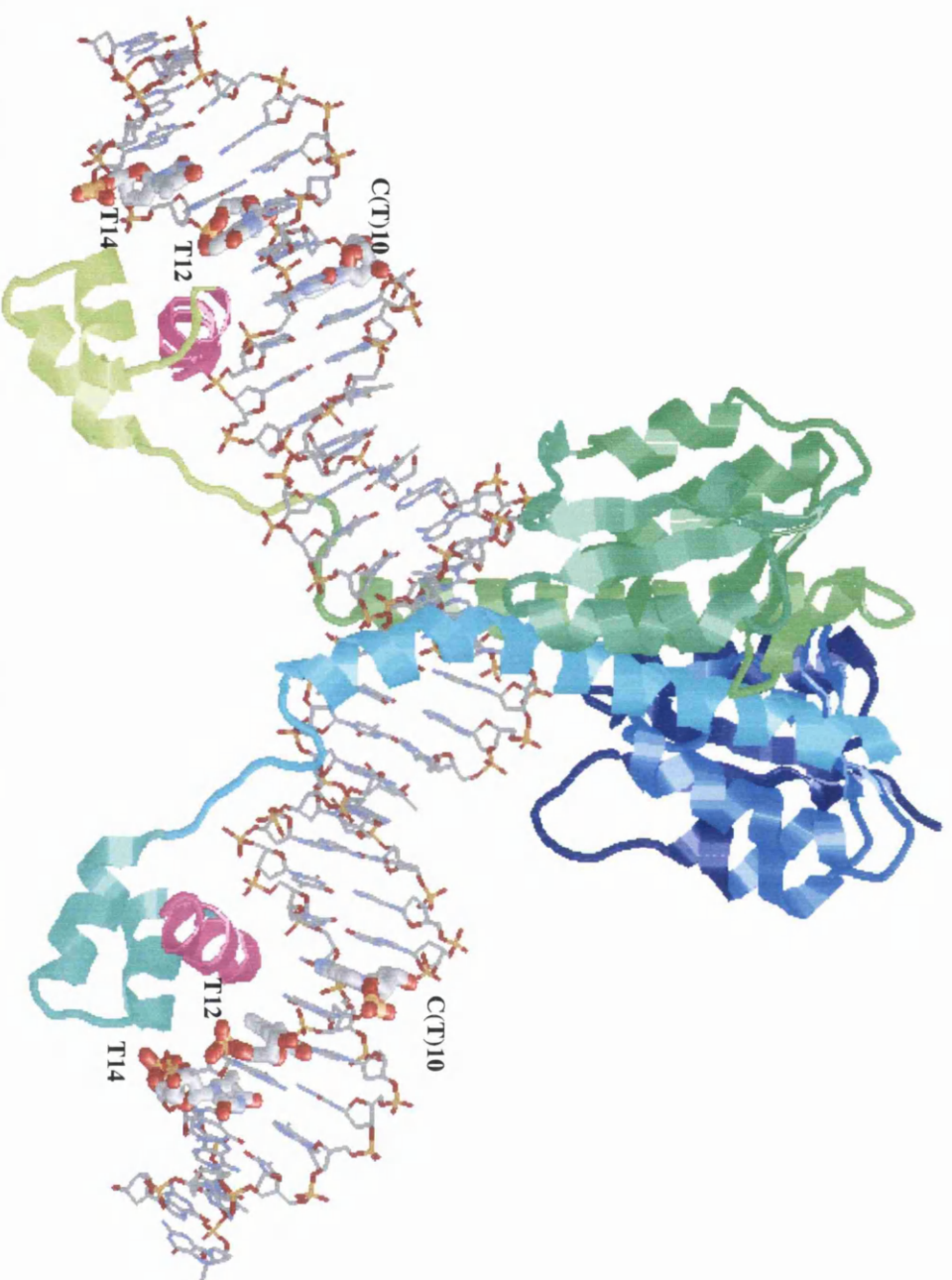


## 4.1 Introduction

The overall aim of the project is to crosslink Tn3 resolvase to its binding site, to gain a greater understanding of the protein-DNA interactions involved. The method chosen for crosslinking is by forming disulfide bonds between a modified nucleotide in the substrate DNA and a cysteine residue in a mutant version of Tn3 resolvase. The resolvase-DNA interactions occurring at site I, the site where the breakage and re-joining of the DNA occurs, will be studied using crosslinking.

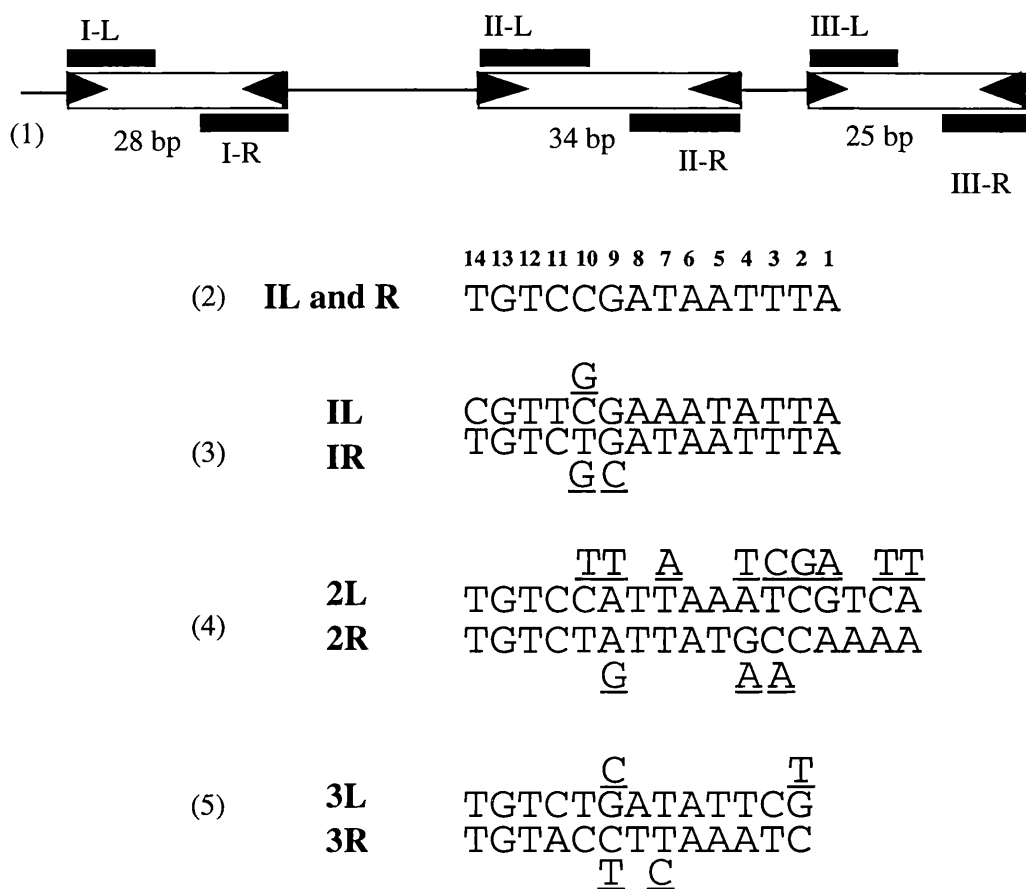
In the  $\gamma\delta$  co-crystal structure (Figure 4.1; Yang and Steitz, 1995) the DNA-binding domain, residues 141-183, makes close contacts with the DNA. The domain binds to the DNA on the opposite side to the N-terminal domain. There is a 9 bp region of highly conserved sequence that constitutes the recognition motifs within *res* sites for binding resolvase. Each *res* site is flanked by a consensus TGT motif (Figure 4.2), and the recognition helix of the resolvase DNA-binding domain interacts with this sequence. Each binding domain of the dimer binds in a single major groove, and has its associated 'hinge' region (residues 140-147) in the adjacent minor groove. The sequence of the symmetrical oligonucleotide used to obtain crystallographic data for the site I resolvase interaction is shown in Figure 4.2.1 and the base at position 10 is a C. The oligonucleotides that were synthesised to contain the modifications have a T at position 10.

The thymines in the binding motif will be replaced by thiol-containing nucleotide analogues (Chapter 2). A cysteine residue must be placed in the recognition helix residues within a reasonable distance from the modified base, so that a disulfide can be made between the two. The requirement for a useful cysteine mutant is that the cysteine thiol must be close enough to the base modification to allow crosslinking. All the residues in the recognition helix that interact with the DNA in the major groove were mutated to cysteines (see Section 4.4). It is possible that mutating this region of the protein may have an adverse effect on the protein-DNA interactions.



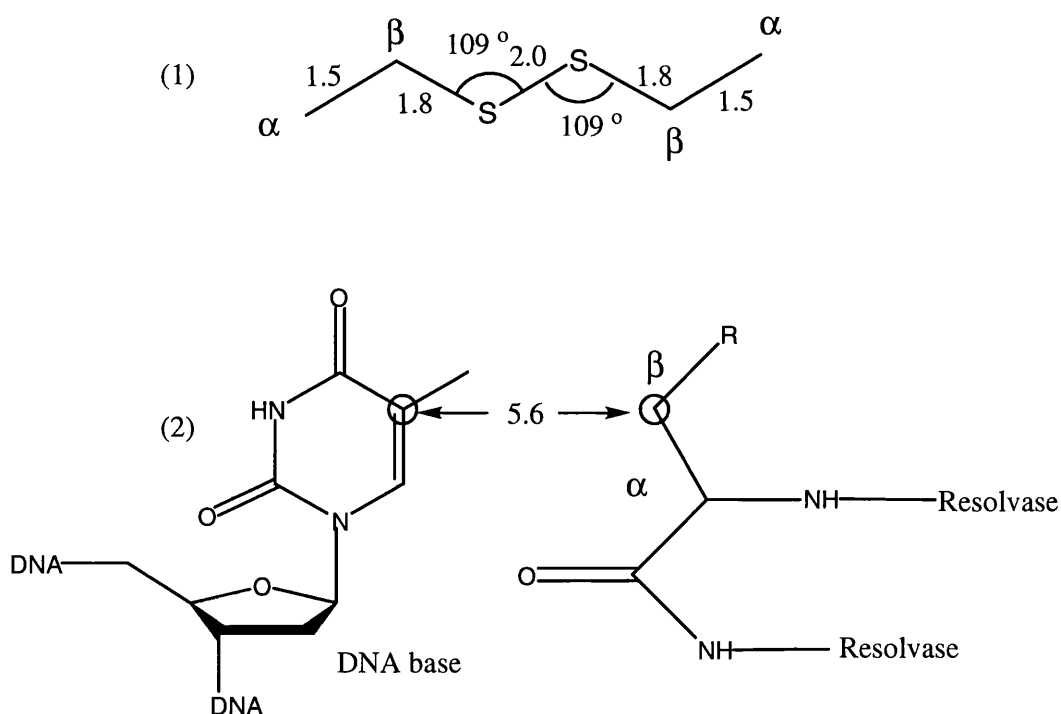
**Figure 4.1 The  $\gamma\delta$  resolvase-site I co-crystal structure with modified areas highlighted**

The bases modified with **SH 17** are numbered and the recognition helices containing the cysteine mutations are highlighted in magenta. Adapted from Yang and Steitz, 1995. The modified oligonucleotides have a thymidine at position 10. In the symmetrical site I used in the crystallographic analysis position 10 is a cytidine.



**Figure 4.2 Sequences bound by Tn3 resolvase**

(1) shows the *res* site. (2) shows the symmetrical site I used in the co-crystal structure (top strand only) of the half site (3, 4 and 5) are the sequences of the natural resolvase binding sites of Tn3 (L = left, R = right). For the right half sites, the homologous (bottom strand) half site sequence is shown. The underlined bases above indicate different bases present in the *res* site of  $\gamma\delta$  resolvase.



**Figure 4.3 Calculation of disulfide bond lengths**

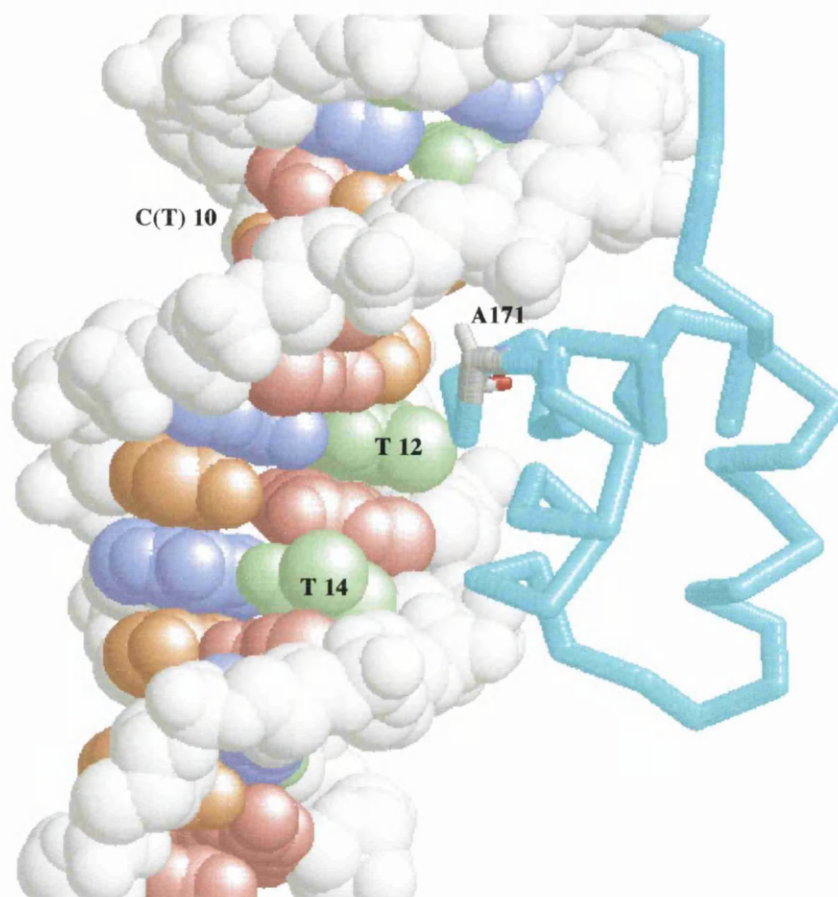
Illustration of how the distances were measured from the co-crystal structure (Yang and Steitz, 1995). (1) shows the bond lengths and angles between the atoms (2) shows the distance between the C-5 of thymidine and the  $\beta$ -carbon atom of the binding domain residue was used. R = A171, R172, S173, Y176 and K177. The true distance between the sulfur atoms, when taking into account the tetrahedral bond angles, shown above will be less than 5.6 Å. However, it can still be predicted that if the two interacting thiols are 5.6 Å apart a disulfide will form between C-5 and the  $\beta$ -carbon atom.

The length of disulfides in the proteins trypsin and chymotrypsin were studied (M. Boocock, pers. comm.). The distance from  $\alpha$ -carbon to  $\alpha$ -carbon in a cystine was 4.2-6.5 Å. For the work here the distances between the interacting thiols was measured from the  $\beta$ -carbon of the amino acid to the C-5 position of the relevant thymine. The length of the sulfur-sulfur bond in disulfides is 2.0 Å and the theoretical maximal distance between  $\beta$ -carbon atoms in a cystine disulfide is 5.6 Å, neglecting the bond angles, charges and interatomic effects (Figure 4.3). The mutants M106C and I110C made in  $\gamma\delta$  resolvase formed dimers *via* disulfide bonds (Hughes *et al.*, 1993). In the co-crystal structure the  $\beta$ -carbon atoms of M106 and I110 were 6.5 Å and 4.2 Å away from the equivalent position in the other dimer. It can be predicted that disulfides in the protein and DNA will form if the distance between the interacting thiols is  $\geq 5.6$  Å.

The residues in the recognition helix A171, R172, S173, T174, V175, Y176 and K177 are all potentially suitable for cysteine mutagenesis and crosslinking as they are in the recognition helix of the DNA-binding domain. Of the residues listed above, T174 and V175 were omitted as they point towards the centre of the DNA-binding domain and are quite far away from the DNA. A171, R172, S173, Y176 and K177 were chosen because they are potentially within a reasonable distance to make a disulfide with a modified base (see Figure 4.4-4.8).

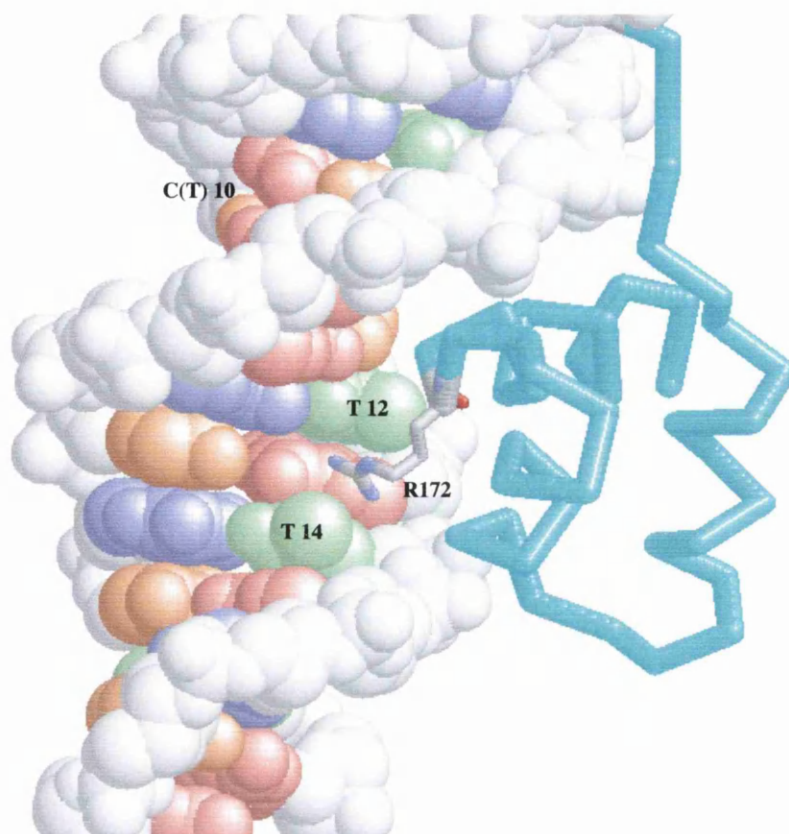
## 4.2 The DNA-binding domain

The last 36 residues of Tn3 resolvase (148-183) constitute the DNA-binding domain. There is also involvement of the 'hinge' region of the protein (residues 141-147), assisting the small domain to bind. The DNA-binding domain of  $\gamma\delta$  resolvase comprises 3  $\alpha$ -helices, residues D149-Q157 (helix 1), S162-T167 (helix 2) and R172-N183 (helix 3). These helices are stabilised by hydrophobic core residues V151, I164 and V175 and by a salt bridge between R148 and E181 in Helices 1 and 3 (Figure 4.9). Helices 2 and 3 form a backbone fold which is similar to the helix-turn-helix motif. Helices 1 and 3 cross with right-handed



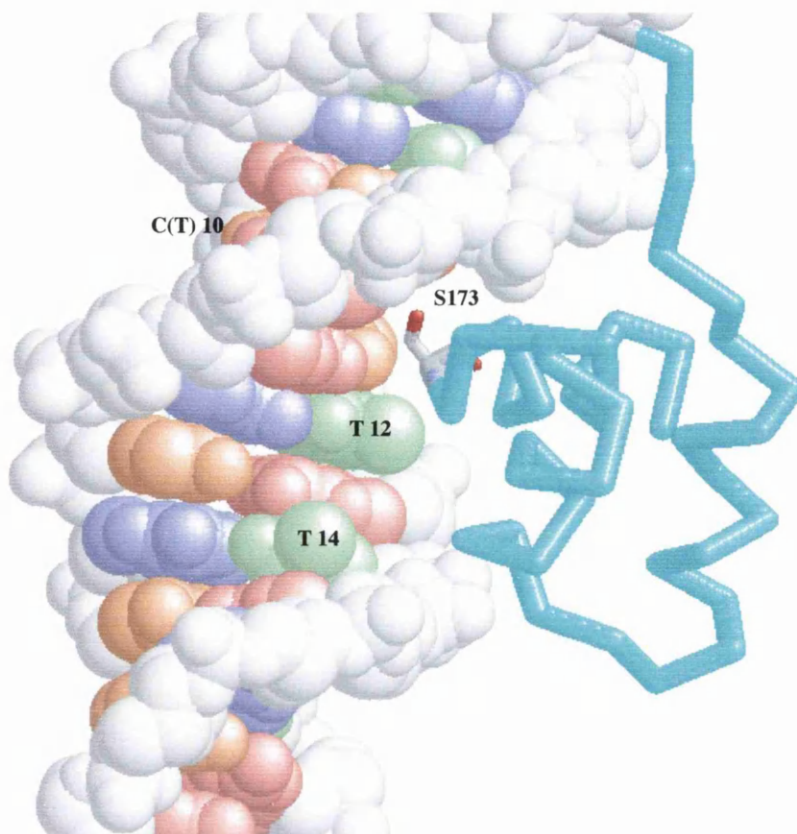
**Figure 4.4 Interactions of A171 with the modified bases**

A171 is shown in CPK (CPK = Corey-Pauling-Koltun colours), the thymidines (green), cytosines (orange), adenines (purple) and guanines (pink). The DNA phosphate backbone is white. The oxygen atom of alanine is red. The bases are numbered as in Figure 4.2. The protein backbone of the C-terminal domain is cyan. The data are derived from Rasmol.



**Figure 4.5 Interactions of R172 with the modified bases**

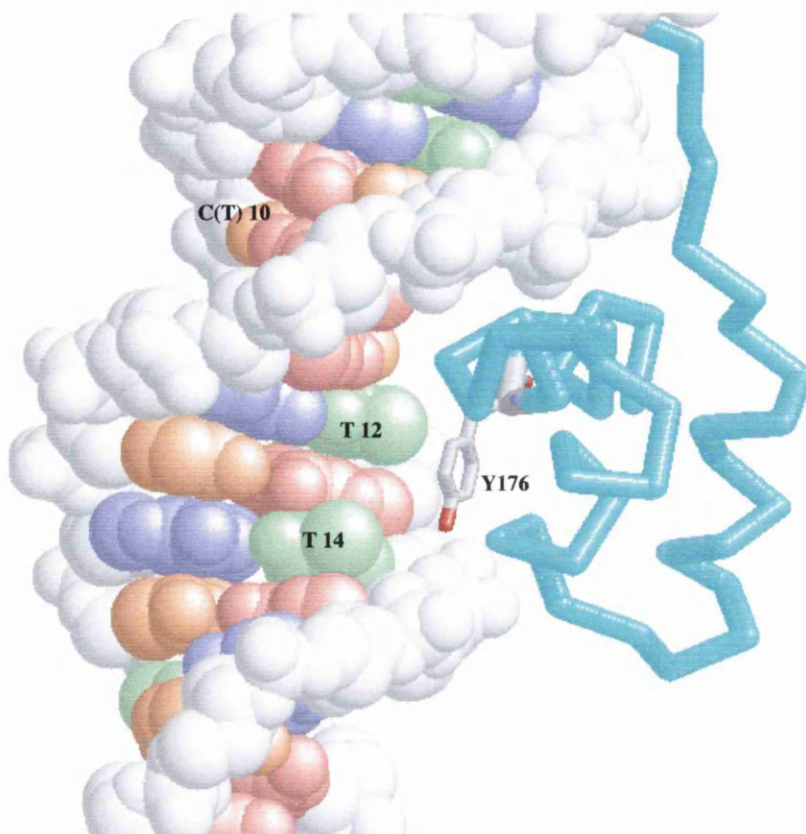
R172C is shown in CPK (CPK = Corey-Pauling-Koltun colours), the thymidines (green), cytosines (orange), adenines (purple) and guanines (pink). The phosphate backbone is white. The oxygen and nitrogen atoms of arginine are red and blue, respectively. The bases are numbered as in Figure 4.2. The protein backbone of the C-terminal domain is cyan. The data are derived from Rasmol.



**Figure 4.6 Interactions of S173 with the modified bases**

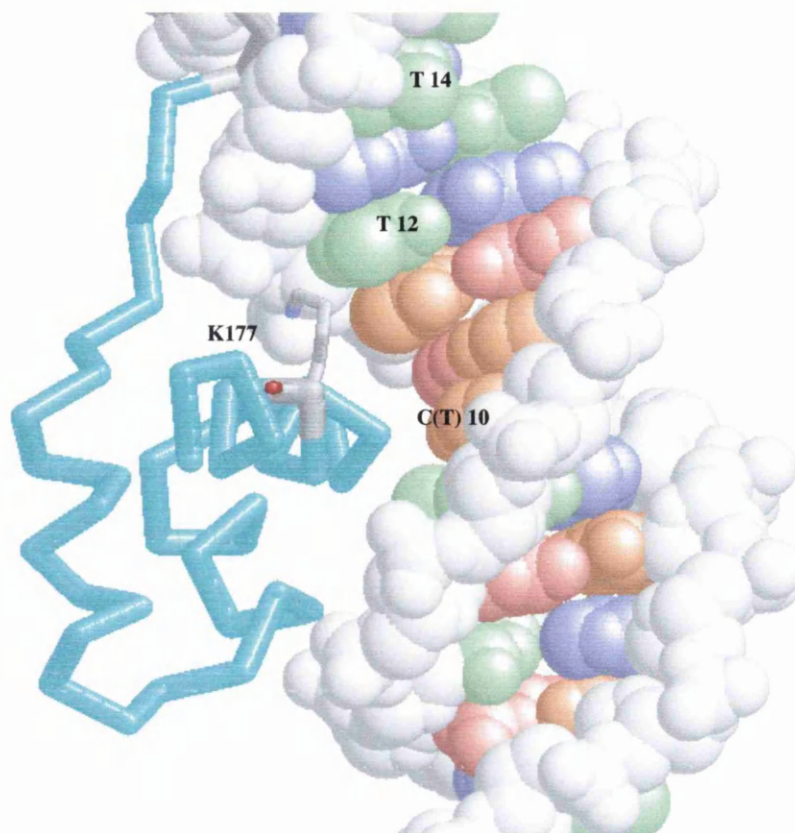
S173C is shown in CPK (CPK = Corey-Pauling-Koltun colours), the thymidines (green), cytosines (orange), adenines (purple) and guanines (pink). The phosphate backbone is white. The oxygen atom of serine is red. The bases are numbered as in Figure 4.2. The protein backbone of the C-terminal domain is cyan. The data are derived from Rasmol.





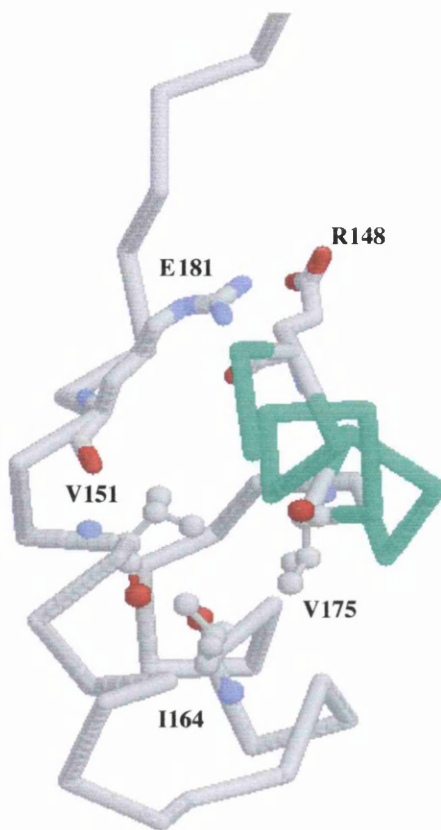
**Figure 4.7 Interactions of Y176 with the modified bases**

Y176C is shown in CPK (CPK = Corey-Pauling-Koltun colours), the thymidines (green), cytosines (orange), adenines (purple) and guanines (pink). The phosphate backbone is white. The oxygen atom of tyrosine is red. The bases are numbered as in Figure 4.2. The protein backbone of the C-terminal domain is cyan. The data are derived from Rasmol.



**Figure 4.8 Interactions of K177 with the modified bases**

K177C is shown in CPK (CPK = Corey-Pauling-Koltun colours), the thymidines (green), cytosines (orange), adenines (purple) and guanines (pink). The phosphate backbone is white. The oxygen and nitrogen atoms of lysine are red and blue, respectively. The bases are numbered as in Figure 4.2. The orientation of the co-crystal structure was changed (compared to Figures 4.4-4.7) in order that the interaction could be seen clearly. The protein backbone of the C-terminal domain is cyan. The data are derived from Rasmol.



**Figure 4.9 The DNA-binding domain of  $\gamma\delta$  resolvase**

The recognition helix is highlighted in green. The highly conserved residues that form the hydrophobic core are labelled in bold. The residues that form a salt bridge between the helices are also labelled in bold.

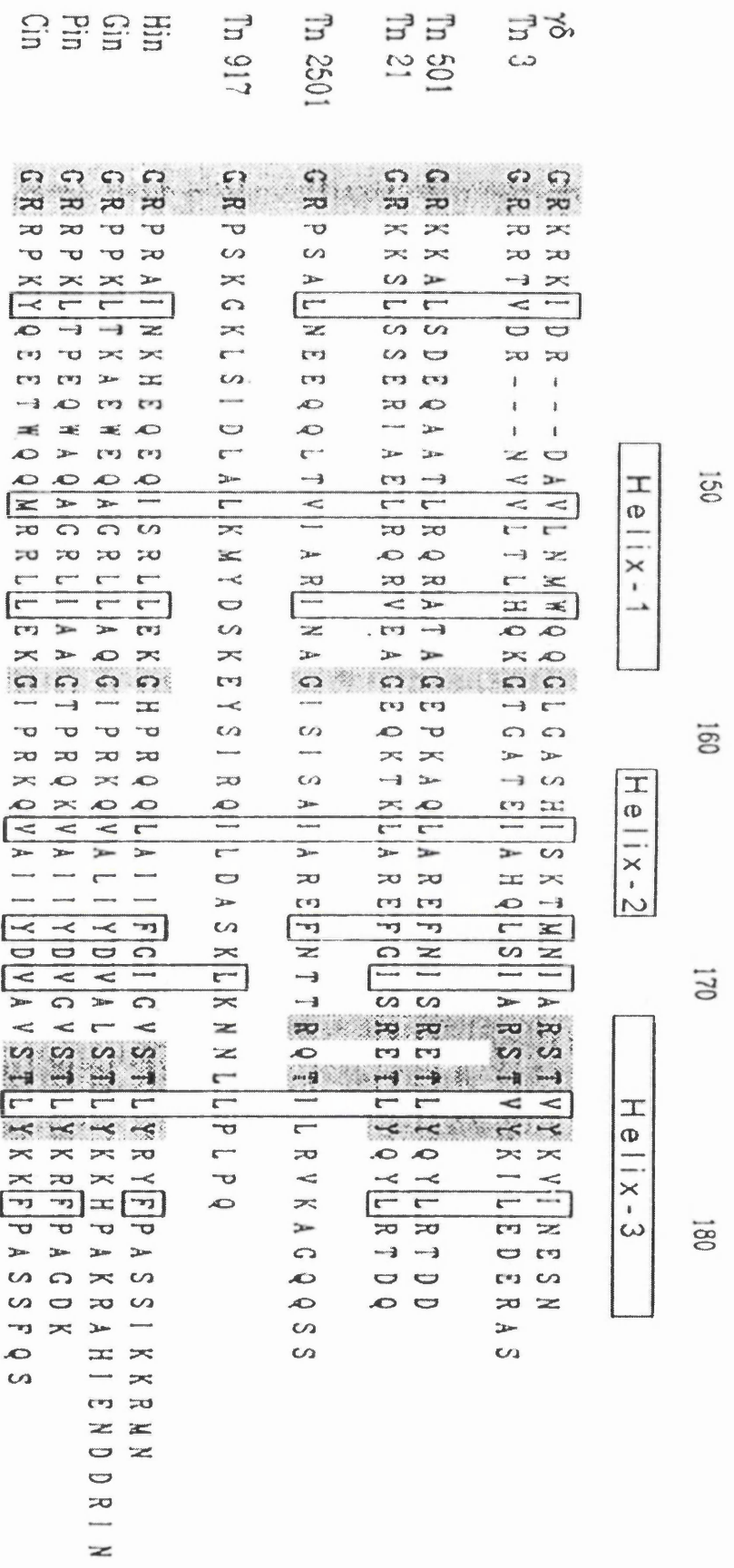


Figure 4.10 Sequence alignments from resolvases and invertases

The shaded regions in the above alignment show areas of conservation within the helix-turn-helix motif for the resolvases and invertases. The boxes represent highly conserved hydrophobic residues.

folding topology (Liu *et al.*, 1994). The helix-turn-helix motif is conserved through most members of the family of site-specific recombinases (Figure 4.10).

Binding of the DNA-binding domain of  $\gamma\delta$  resolvase has been studied using DNase I footprinting (Abdel-Meguid *et al.*, 1984) and ethylation interference experiments (Rimphanitchayakit *et al.*, 1989).  $\gamma\delta$  resolvase was cleaved by chymotryptic digestion, isolating the N and C-terminal domains. The intact DNA-binding domain interacts at *res* subsites similarly to intact resolvase. A minimum 12 bp binding sequence motif at both ends of site I was protected by the DNA-binding domain, indicating that this is the region of site I where the C-terminal domain binds. There was an additional area of protection at the centre of site I with the intact protein, on the opposite side to the regions protected by the DNA-binding domain, indicating that the catalytic and binding domains are on opposite sides of the DNA. The ethylation experiments confirmed the footprinting results. A strong inhibition of binding was seen at ethylated phosphates at positions IR-G13 and IR-T12 and an additional region of protection at positions IR-T8 and IR-A7 towards the centre of site I.

The sequence requirements for the binding of the C-terminal domain of  $\gamma\delta$  resolvase to its recognition site were determined by saturation mutagenesis (Rimphanitchayakit and Grindley, 1990). Mutations were made at all positions in site I. The regions that were most affected by changing the bases could be divided into two regions; positions IR-T14-8 at the ends where the DNA-binding domain binds, and positions IR-C9 and IR-G10 towards the centre of the site, where the intact protein binds. The numbering of the bases in site I is explained in Figure 1.2. The substitution IR-C4 to IR-G4 in site I increased the binding affinity of  $\gamma\delta$  resolvase for the site I compared to the wild-type site.

### 4.3 Cysteine Mutants of Tn3 resolvase

Cysteine mutants in Tn3 and  $\gamma\delta$  resolvases have been created previously, for mechanistic studies (Hatfull and Grindley, 1986), crystallographic studies (Yang and Steitz, 1995), crosslinking between subunits (Hughes *et al.*, 1993; Murley and Grindley, 1998; Weinwieser, 2001; MacDonald, 1999) and attachment of chemical modifications (MacDonald, 1999; Mazzarelli *et al.*, 1993).

The cysteine mutant T162C was made for crosslinking analysis (MacDonald, 1999) *via* the formation of a disulfide bond with a phosphorothioate-containing site I. Position 162 is in close contact with the phosphodiester at IR-T14. This phosphate was replaced with a phosphorothioate, which should form a disulfide with T162C under oxidising conditions. However no crosslinked complexes were obtained using this strategy, as the sulfur of the phosphorothioate was not reactive enough to form disulfides with the cysteine of T162C. The work here that has DNA bases containing thiols (as opposed to the DNA backbone) should be more effective at forming disulfide crosslinked complexes. There have been mutations made in  $\gamma\delta$  resolvase recognition helix 3. These are R172L, R172G, S173G, and Y176C.

### 4.4 Isolation of Novel DNA-binding domain mutants

All the positions chosen to be mutated, except A171, form base-specific contacts with the DNA. Within the DNA substrates there is a region of sequence at positions 13-10 (Figure 4.2), GTCC, that forms an essential portion of the binding site where most of the chosen residues form interactions. During the co-crystallisation studies (Yang and Steitz, 1995), it was observed that five out of the six of recognition sequences at the ends of the three subsites of *res* had at least one IR-G10 and IR-G11 (Figure 4.2) in the site I, indicating that a guanine at these positions is important for recognition by the DNA-binding domain.

A171 is positioned outside the consensus TGT motif, so is not essential for binding, but its close proximity to the DNA backbone makes it a good candidate

for crosslinking. The  $\beta$  carbon of A171 is 11-13 Å away from the three thymidines in site I that will be modified with thiol groups, longer than the calculated value for a disulfide bond (Figure 4.4). However the interacting sulfur atoms in A171C and the thiols in the DNA may come closer during the reaction.

R172 has an electrostatic interaction with O6 and N7 of IR-G13, with a further hydrophobic interaction with the methyl group of IR-T14, the first T in the consensus TGT (Figure 4.5). From Table 4.1 the distance between the C-5 atom of T14 and the  $\beta$  carbon of R172 is 7.91 Å, closer to the ideal value for a disulfide, making the formation of a disulfide between the two thiols plausible. The  $\beta$  carbon atom of R172 is further away from T12 and C(T)10 at 8.28 and 11.75 Å respectively, perhaps still close enough for disulfide formation to occur.

In  $\gamma\delta$  resolvase, the arginine at position 172 has been mutated to leucine (R172L) and also to glycine (R172G). A screening experiment found that the R172L mutant binds selectively to complete *res* sites with the site I sequences mutated at IR-T14, IR-T13 and IR-T12 (Grindley, 1993; cf IR-T14, IR-G13 and IR-T12 in wild-type binding sites), and was selected for this property. The R172L mutant was isolated and purified, and was found to have activity similar to wild-type  $\gamma\delta$  resolvase, although it had a weaker binding interaction with wild-type *res* sites. This preferential binding of R172L has been used extensively in targeting experiments (Murley *et al.*, 1998, Wienwieser, 2001) designed to direct resolvase monomers to specific sites within *res*. Another mutant at this position, R172G, was found to have lost all recombinational activity.

The hydroxyl group of S173 is involved in hydrogen bonding with the O-6 atom of IL-G10 (Figure 4.6). This residue of  $\gamma\delta$  resolvase was mutated to glycine and purified. S173G displayed weaker binding activity than the wild-type protein (Grindley, 1994), but retained recombination activity. The  $\beta$  carbon atom of S173 is closest to T12 and C(T)10 from Table 4.1, being around 6 Å from the C-5 of the thymidines. Disulfide formation should be possible between these positions as

they are close enough. S173 interacts with the DNA towards the centre of the site, therefore it is 10.55 Å away from the flanking T14.

The residue Y176 interacts closely with the IR-T12 in the TGT motif and is 5.35 Å away from it (Figure 4.7). This interaction is the closest of all measured in Table 4.1 and is close to the ideal distance between the β-carbons in a disulfide. Therefore it can be predicted that a thiol at the C-5 of thymidine and the an equivalent thiol on the β carbon atom of Y176 should form a disulfide. The residue associates with the methyl group of thymine *via* its phenolic aromatic ring and also the β carbon of the amino acid side chain. γδ resolvase Y176C was found to be inactive *in vivo*.

K177 makes a hydrophobic interaction with the methyl group of IR-T8 and is approximately 9 Å away from T12 and T14 in the DNA backbone (Figure 4.8), making the concept of disulfide bond formation as mentioned above, as possibility.

From the co-crystal structure (Yang and Steitz, 1995) using the Rasmol program, the distances for crosslinking were measured. The distances were from the β-carbon atom of the amino acid residue to the C-5 of thymidine (Figure 4.3) and are measured in angstroms (Å).



**Table 4.1**

proteins→ T analogue↓	A171	R172	S173	Y176	K177
<b>T</b> GTCT (T14)	Cβ---C-5 <u>12.87</u> (13.07)	Cβ---C-5 <u>7.71</u> (7.91)	Cβ---C-5 11.32 (10.55)	Cβ---C-5 <u>9.81</u> (10.19)	Cβ---C-5 <u>14.67</u> (14.99)
TG <b>T</b> CT (T12)	Cβ---C-5 <u>10.49</u> (10.91)	Cβ---C-5 <u>7.12</u> (8.28)	Cβ---C-5 <u>6.31</u> (6.24)	Cβ---C-5 <u>5.14</u> (5.35)	Cβ---C-5 <u>8.48</u> (9.01)
TGT <b>C</b> T (C10)	Cβ---C-5 <u>10.55</u> (10.59)	Cβ---C-5 <u>11.29</u> (11.75)	Cβ---C-5 <u>6.35</u> (6.71)	Cβ---C-5 9.94 (9.59)	Cβ---C-5 <u>8.18</u> (9.20)

The numbers refer to the distances from the  $\beta$ -carbon-C-5 methyl group in the target bases and amino acids in the co-crystal structure from the Rasmol programme. The numbers that are unbracketed represent the distances from one side of the structure, the bracketed numbers represent the corresponding measurements made from the other side of the symmetrical site I. The underlined numbers indicate that although the same distances were measured, one side of the site I-resolvase interaction in most cases is closer.

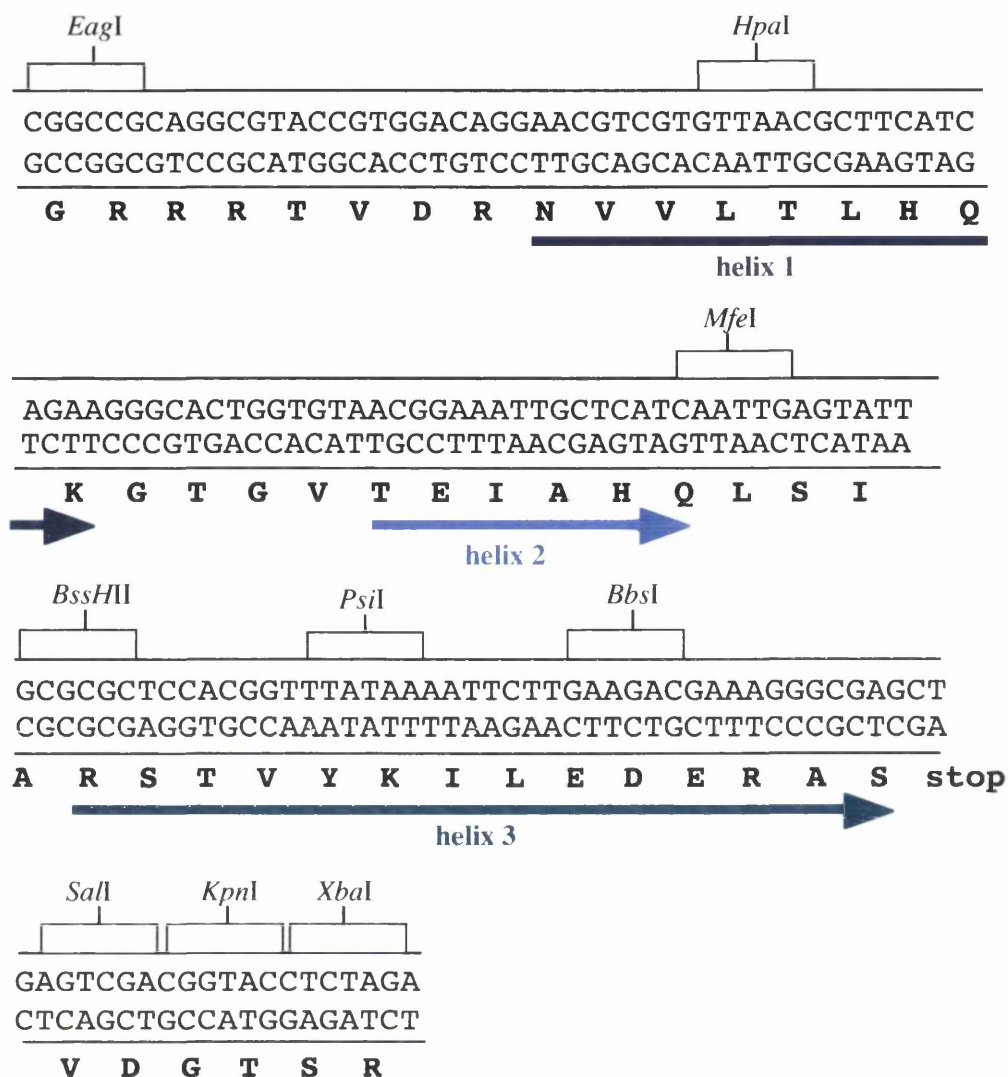
From Figure 4.3 a distance of  $< 5.6 \text{ \AA}$  between the C-5 of the base and the  $\beta$ -carbon should be acceptable for disulfide formation between the two thiols. Distances between the  $\beta$ -carbon of IL-T14 and C-5 range between 7.71-14.67  $\text{\AA}$ . The IL-T12 distances range between 5.14-10.49  $\text{\AA}$  and for the IL-C(T)10 position between 6.35-11.29  $\text{\AA}$ . It can predicted that disulfide bond formation should occur with modification at IL-T12 as it is close enough in some cases. However, the measured distances are taken from the co-crystal structure, and the atoms may come physically closer during the recombination reaction. The modified oligonucleotides have a thymidine at position 10. In the symmetrical site I used in the crystallographic analysis, position 10 is a cytidine.

#### 4.5 The construction of plasmids encoding cysteine mutants

The DNA-binding domain open reading frame of Tn3 resolvase is shown in Figure 4.11. The three helices are indicated below the sequence; the recognition helix is the longest of the three. The plasmid pMA6211 contains the resolvase open reading frame between sites *NdeI*-*Asp718*, carrying the T162C mutation (MacDonald, 1999). This plasmid was chosen because it contains restriction sites *HpaI* and *BssHI* in the last third of the resolvase reading frame. These sites were used to clone *HpaI*-*BssHII* oligonucleotides, which remove the T162C mutation, into the equivalently cut pMA6211.

The oligonucleotides, top strand 56 nucleotides (nt) and bottom strand 60 nt, (Figure 4.12.1) cloned between these sites carried an additional *MfeI* site. The new plasmid, pMS81, contains a unique *MfeI* site and two *PsiI* sites and encodes wild-type resolvase. To facilitate cloning of the oligonucleotides containing the cysteine codons, it was necessary to remove the other *PsiI* site. This was achieved by a full *SspI* (three *SspI* sites) digest and a partial *PsiI* digest. The 4.9 kb fragment was re-circularised yielding pMS82, which contains a unique *PsiI* site. The mutant oligonucleotides that carried the five mutant codons (Figure 4.12.2) at positions A171, R172, S173, Y176 and K177 in the resolvase reading frame (Figure 4.11.1), were cloned from the *MfeI*-*PsiI* sites of pMS82 for A171C, R172C, S173C and Y176C, and from *PsiI* to *SalI* for K177C, yielding the plasmids pMS89(A171C), pMS90(R172C), pMS91(S173C), pMS92(Y176C) and pMS93(K177C).

However, these constructs would only transform the strain DH5 $\alpha$ . The reasons for this are unknown, but are probably due to the deletion of the *PsiI*-*SspI* fragment near the *bla* promoter. This meant that these plasmids could not be used in the *in vivo* assay (Section 4.7). The C-terminal sections of the reading frames were therefore excised by *EagI*-*Asp718* digestion and replaced back into pAT6 $\Delta$  (MacDonald, 1999), yielding plasmids pMS95(A171C), pMS96(R172C), pMS97(S173C), pMS98(Y176C), pMS99(K177C) and pMS100 (wild-type).



**Figure 4.11** The sequence of the C-terminal domain of Tn3 resolvase

The three helices are shown by the arrows. The restriction sites used for cloning and analysis are shown. The amino acid codons are written in bold beneath. The sequences of three  $\alpha$  helices of the DNA binding domain are shown beneath the codons as blue arrows.

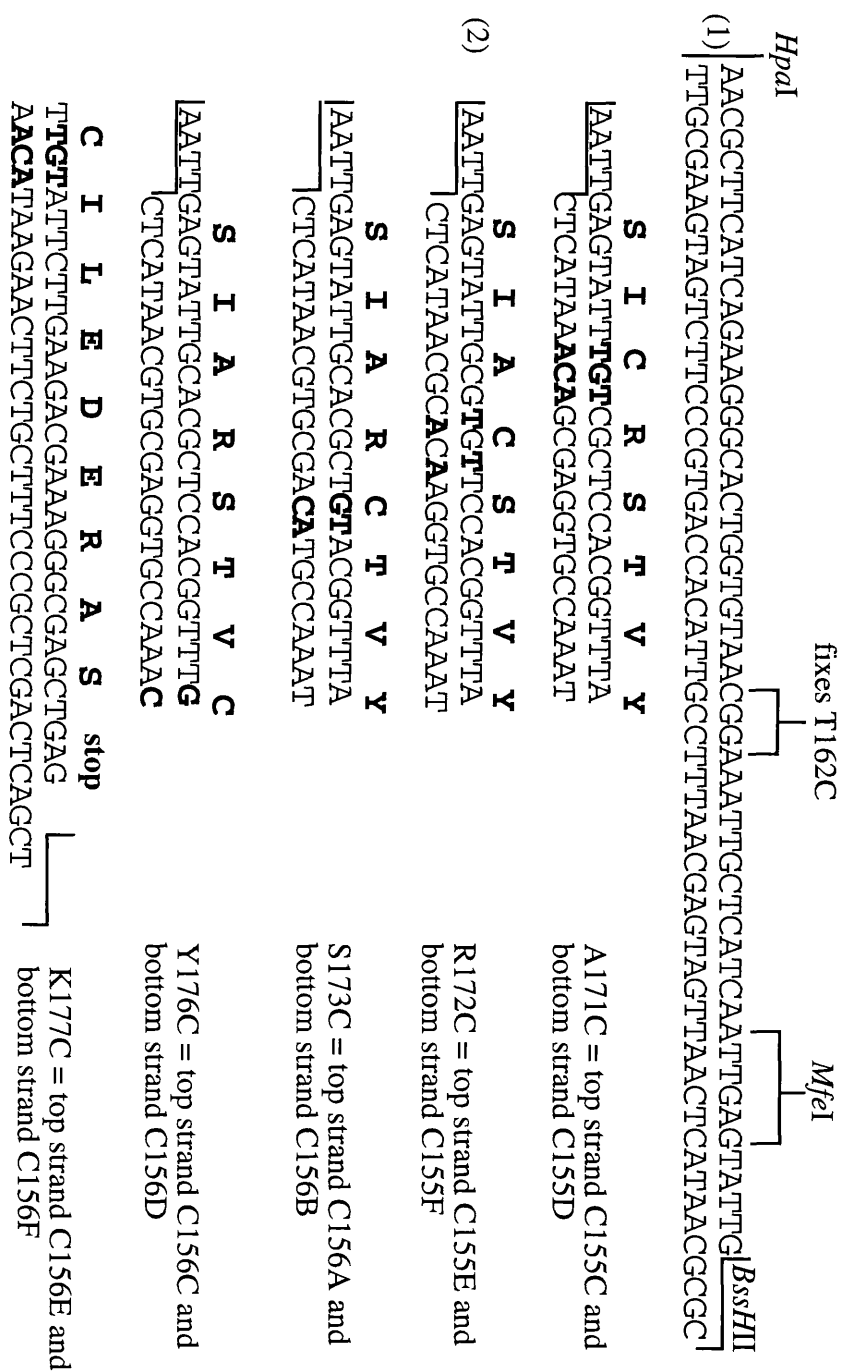


Figure 4.12 Illustration of the oligonucleotides used for cloning the novel cysteine mutants

(1) shows the sequence of the *HpaI*-*BssHII* linker oligonucleotide used in the first cloning step. The cloning of this oligonucleotide into the equivalently cut pM fixed the T162C mutation and introduced the new restriction site *MfeI* that was used to clone the smaller oligonucleotides shown in (2). Positions of the cysteine codons are shown in bold in the oligonucleotides used to clone into the template plasmid pMS82.

pMS100 is illustrated in Figure 4.13. These plasmids could be tested *in vivo* (Section 4.7). Cloning was carried out by M. Burke.

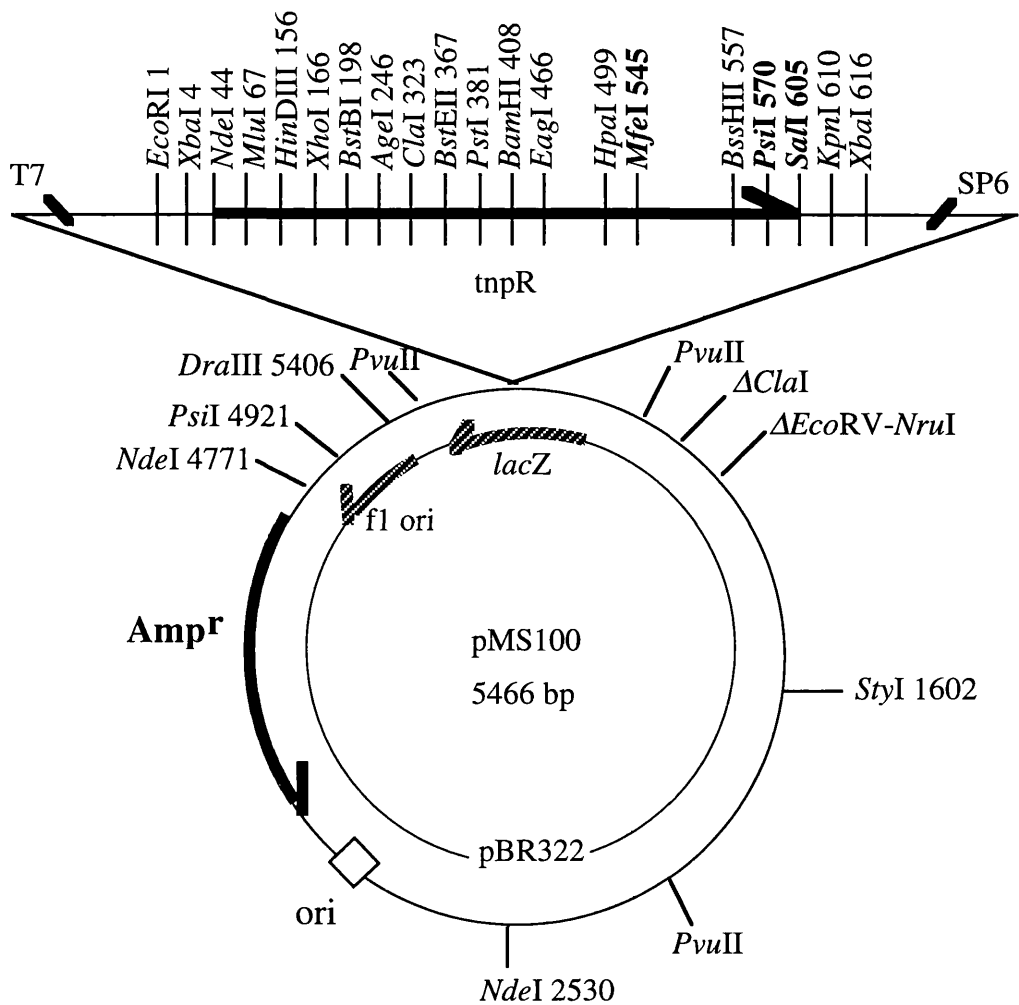
#### 4.6 Cloning of C-terminal sequences into the expression plasmid

pALY15 (6693 bp) is a pSA1101-based (S. Rowland) overexpression plasmid that carries the T162C mutant resolvase open reading frame (MacDonald, 1999). pALY15 was digested with *Asp*718 and *Eag*I to remove the DNA that encodes the C-terminal domain of the T162C mutant. The plasmids carrying the DNA-binding domain mutants pMS95-pMS99, were also digested with *Asp*718 and *Eag*I, thus releasing DNA-binding domain fragments of the mutants A171C to K177C. The large fragments from the pALY15 digest (6548 bp) and the small fragments (145 bp) from the pMS100-derived plasmids were isolated using agarose (Section 3.21) and acrylamide gel electrophoresis (Section 3.24.2), respectively.

The fragments were ligated, and the ligation products were used to transform DS941. The clones were checked by single colony analysis and restriction enzyme digests to ensure that the reading frames were correctly constructed. The absence of certain restriction sites in the clones that arose from the initial cloning of the mutant oligonucleotides, was also confirmed; i.e., the oligonucleotides for A171C, R172C, S173C and Y176C were designed to remove the *Bss*HIII site in the plasmids, and the K177C oligonucleotide should delete the *Bbs*I site. The Y176C and K177C mutations also destroy the *Psi*I site in this region, because the codons overlap with the site. The new C-terminal sequences are illustrated in bold letters in the oligonucleotides used for the cloning (Figure 4.12.2).

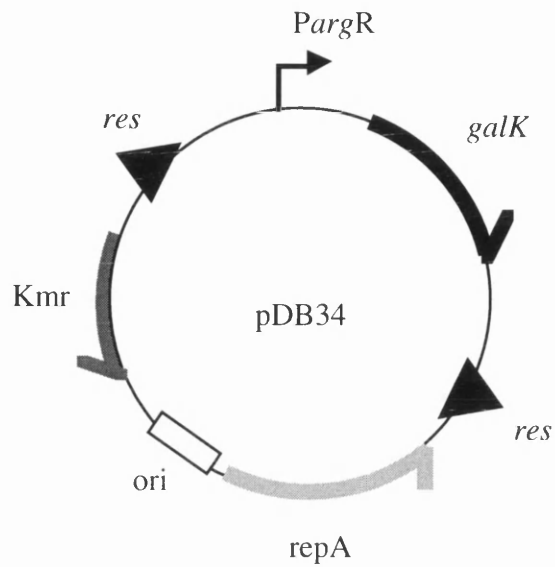
#### 4.7 *In vivo* recombination assay

The plasmids containing the mutant resolvase were transformed into DS941 containing one of three different plasmids. pDB34 contains two full *res* sites in direct repeat (Figure 4.14). Competent cells of DS941 carrying the test plasmid pDB34 were prepared using kanamycin selection for cells containing the



**Figure 4.13 Illustration of pMS100**

The diagram shows the plasmid used for the *in vivo* screen. This is the wild-type version of the plasmids pMS95-99 that encode the cysteine mutants.



**Figure 4.14 Diagram of resolvase *in vivo* screening plasmid pDB34**

The plasmid used for the *in vivo* screening of resolvase cysteine mutants (Arnold *et al.*, 1999). The origin of replication and *repA* gene for the replication initiator protein are from pSC101. The *galK* gene is transcribed from *PargR*, the promoter of the *E.coli argR* gene.

substrate. These cells were then transformed with pMS95-99 encoding the cysteine mutants and plated onto media containing galactose (See Section 3.37). Any resolution of pDB34 by the mutants results in loss of the *galK* gene from pDB34, and so white colonies grow on the plates.

The results of the *in vivo* testing for recombination are shown in Figure 4.15. Three out of the five cysteine mutants show *in vivo* recombination activity. The recombinationally active proteins are A171C, S173C and K177C, indicated by the presence of white colonies on the MacConkey agar. The mutants that are unable to carry out recombination in this assay, i.e. R172C and Y176C, are likely to be deficient in binding activity.

#### **4.8 Expression of the mutants in the strain BL21(DE3)pLysS**

For details see Section 3.34.

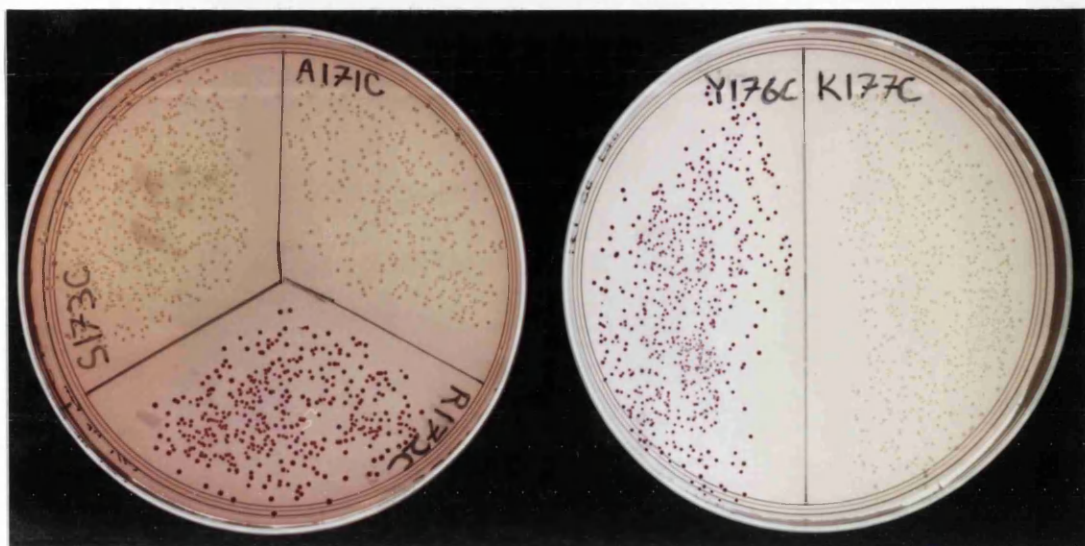
Derivatives of the *E.coli* strain BL21 were used to express the Tn3 resolvase cysteine mutants. The BL21(DE3) strain carries a stable bacteriophage  $\lambda$  lysogen derivative ( $\lambda$  DE3), in the chromosome of BL21, which contains the gene for T7 polymerase under the control of a *lacUV* promoter. The *lacUV* promoter is bound by *lac* repressor, and is inducible by isopropyl- $\beta$ -D-thiogalactopyranose (IPTG).

T7 lysozyme is an inhibitor of T7 polymerase, so it reduces the potentially damaging effects of transcription before induction (Studier *et al.*, 1990). The strain BL21(DE3)pLysS contains the gene encoding T7 lysozyme on a plasmid conferring chloramphenicol resistance. The low level of lysozyme production by this strain has little effect on the induction of target genes, so it is very efficient at producing high levels of target protein in the presence of IPTG.

#### **4.9 Purification of cysteine resolvases**

For experimental details see Section 3.35.





**Figure 4.15 *In vivo* recombination of pDB34 by A171C, R172C, S173C, Y176C and K177C**

The left panels show A171C, R172C and S173C, and the right panel shows Y176C and K177C *in vivo* recombination.

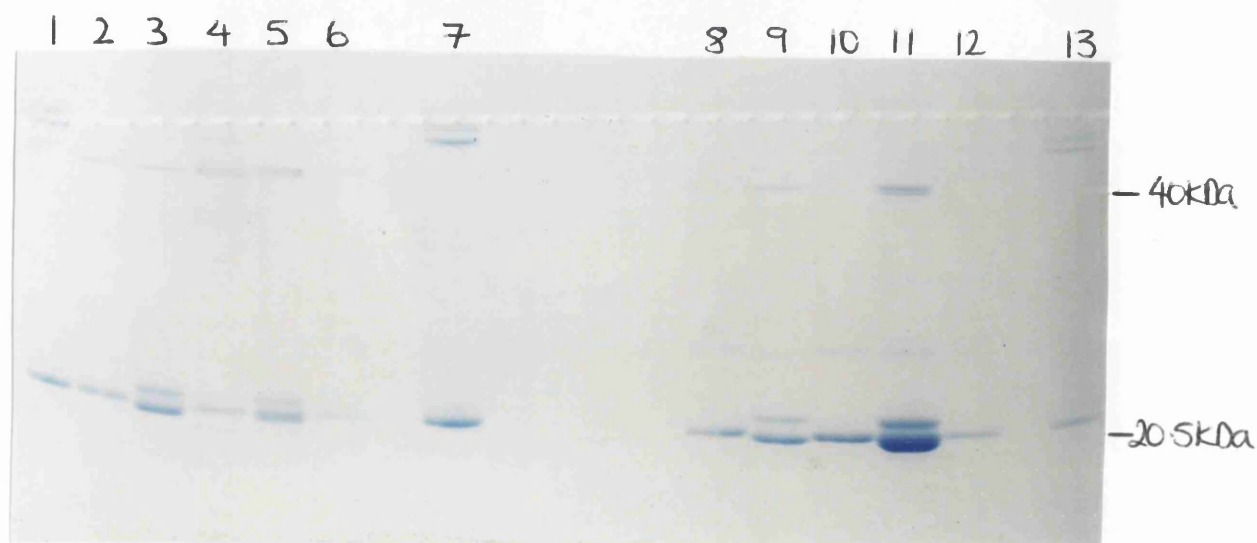
Red colonies = no resolution and white colonies = resolution.

The method of purification used is a denaturing preparation that exploits the solubility of resolvase in buffers containing varying concentrations of salt. At high salt concentrations resolvase is solubilised, whereas in buffers containing low salt concentrations resolvase is insoluble and so can be precipitated. Initially the cells are disrupted using sonication, ensuring that the temperature is kept low to reduce protein degradation. After disruption the cells are spun, which pellets the cell debris and the resolvase. After a series of washings at low salt, the cell debris and other contaminant proteins are removed by centrifugation. The final wash at 2 M NaCl solubilises the resolvase, and following centrifugation, the remaining proteins are pelleted, while the resolvase remains in solution. To ensure that all the resolvase is recovered, an additional 2 M salt/6 M urea step ensured solubilisation of the resolvase in the pellet. After checking the purification using Laemmli gels (Section 3.24.1), the denatured resolvase-containing fractions were dialysed against 2M salt buffer to allow the protein to refold. The resolvase was precipitated by dialysis against a low salt buffer. The precipitate was collected and dissolved in buffer containing 6 M urea.

The resolvase fractions were purified further using ion exchange chromatography. Two cation exchange columns were run, the first with a linear gradient of 25-250 mM NaCl for 25 minutes then 250-1000 mM NaCl for a further 25 minutes. The second column used a linear gradient of 25-1000 mM NaCl, run over 1 hour to achieve better separation of the proteins. The cysteine mutants eluted between 280-300 mM NaCl.

#### **4.10 Concentration of cysteine mutants**

The standard A113C was used to determine the concentration of the cysteine mutants as it had been calibrated with R17f33 (48  $\mu$ M; M. Boocock). The concentration of A113C is 40  $\mu$ M. R17f33 is an HPLC fraction of wild-type resolvase that has been determined by amino acid analysis to have a concentration of 48  $\mu$ M, and is used as the lab benchmark standard for determining the concentration of resolvase solutions.



**Figure 4.16 Concentration of cysteine mutants**

Lane	Volume loaded (μl)	Cysteine resolvase	Concentration (μM)	Concentration (mg/ml)
1	0.5	A113C	20	0.4
2	5	A171C	-	-
3	5	R172C	-	-
4	5	S173C	-	-
5	5	Y176C	-	-
6	5	K177C	-	-
7	1	A113C	40	0.8
8	5	A171C + 5 mM DTT	3	0.06
9	5	R172C + 5 mM DTT	8	0.16
10	5	S173C + 5 mM DTT	8	0.16
11	5	Y176C + 5 mM DTT	40	0.8
12	5	K177C + 5 mM DTT	2	0.04
13	0.25	A113C	10	0.2

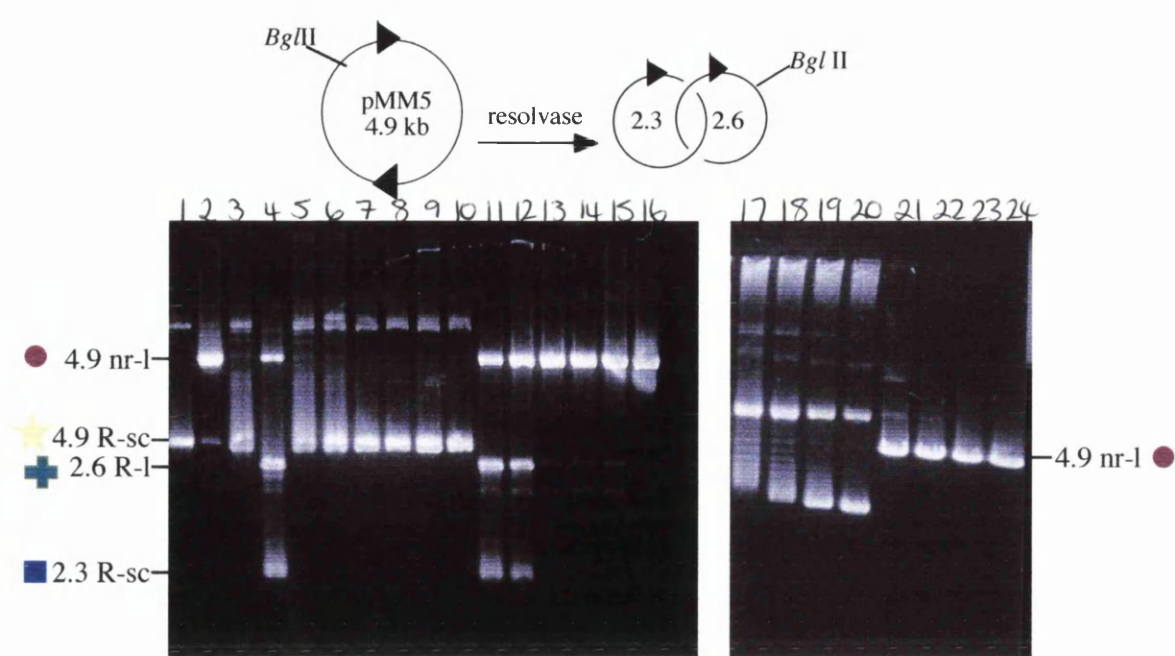
The left half of the gel (Figure 4.16) shows samples of the protein stock solutions stored at 1 mM DTT. The higher band shows that some oxidation has occurred. The geometry of the protein is such that these interactions are unlikely, although prolonged storage must have resulted in disulfides being formed between cysteine residues in the binding domain. The right half represents stock solutions that have had 5 mM DTT added to them. To calculate the concentration, the right hand samples were used. The extra band in some of the samples is an unknown contaminant that elutes at the same salt concentration as resolvase (S.Weinwieser, pers. comm.).

#### **4.11 *In vitro* resolution of pMM5 using novel cysteine mutants**

The five cysteine mutants were tested for their ability to carry out the resolution of pMM5. This assay was used to determine whether or not these novel binding domain mutants were active in recombination under standard *in vitro* conditions. The substrate plasmid is a supercoiled molecule with two *res* sites. The main product from the reaction with resolvase is a supercoiled catenane. After the resolvase was inactivated by heat treatment the products were digested with specific restriction endonucleases, here *Bgl*III was used. The possible products from the *Bgl*III digest are linearised starting plasmid (unresolved product), supercoiled circle (one in catenane that lacks *Bgl*III site) and linear DNA (from circle in catenane that contained the *Bgl*III site).

Figure 4.17 shows the resolution of pMM5 by A171C and R172C. The results of the resolution of pMM5 with S173C, Y176C and K177C are shown in Figure 4.17.

A171C, S173C and K177C all show recombination activity. The concentration of wild-type resolvase was 12  $\mu$ M. A171C was active at 3  $\mu$ M and 1.5  $\mu$ M (Figure 4.17 lanes 11 and 12). S173C was active at 8  $\mu$ M and 4  $\mu$ M (Figure 4.18 lanes 7 and 8) and K177C was active at 2  $\mu$ M (Figure 4.18 lane 22). The mutants A171C and S173C are as active as wild-type resolvase. K177C was active, although the

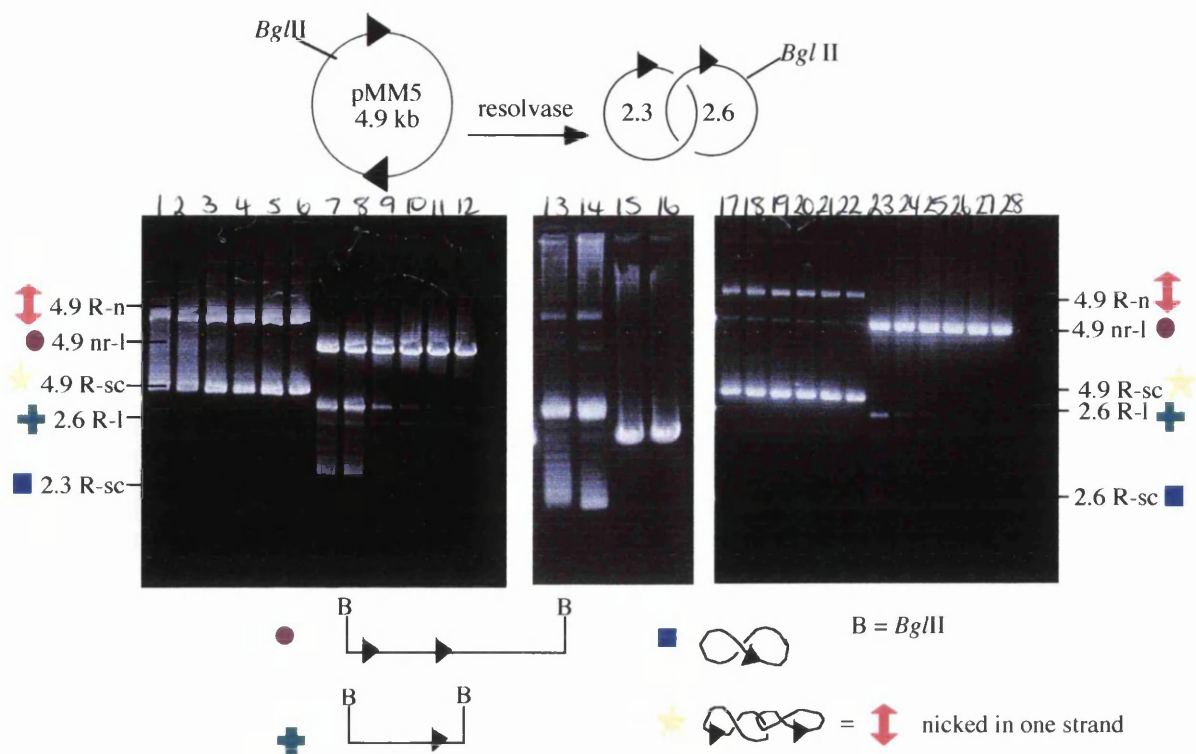


**Figure 4.17 Resolution of pMM5 with A171C and R172C**

R = recombinant, l = linear, sc = supercoiled, nr = non-recombinant

The reactions were carried out using C8.2 buffer and 0.4 µg/lane of DNA and were incubated with 1/10th volume of diluted protein solution at room temperature. The reaction mixture was adjusted to 1 mM DTT prior to addition of the cysteine mutants. The reaction was stopped by heating to 70 °C for 5 minutes, then analysed by restriction enzyme digestion.

- Lanes 1 - No resolvase (dilution buffer), 1 hour  
 2 - as lane 1 + *Bgl*II  
 3 - 16-fold dilution of wild-type resolvase, 1 hour  
 4 - as lane 3 + *Bgl*II  
 5 - Undiluted A171C resolvase  
 6 - 2-fold dilution of A171C resolvase, 1 hour  
 7 - 4-fold dilution of A171C resolvase, 1 hour  
 8 - 8-fold dilution of A171C resolvase, 1 hour  
 9 - 16-fold dilution of A171C resolvase, 1 hour  
 10 - 32-fold dilution of A171C resolvase, 1 hour  
 11-16 as lanes 5-10 + *Bgl*II  
 17 - Undiluted R172C resolvase, 24 hours  
 18 - 2-fold dilution of R172C resolvase, 24 hours  
 19 - 4-fold dilution of R172C resolvase, 24 hours  
 20 - 8-fold dilution of R172C resolvase, 24 hours  
 21-24 as lanes 17-20 + *Bgl*II



**Figure 4.18 Resolution of pMM5 with S173C, Y176C and K177C**

R= recombinant, l = linear, sc = supercoiled, nr = non-recombinant, n = nicked.

The reactions were carried out using C8.2 buffer and 0.4 µg/lane of DNA and were incubated with 1/10th volume of diluted protein solution at room temperature. The reaction mixture was adjusted to 1 mM DTT prior to addition of the cysteine mutants. The reaction was stopped by heating to 70 °C for 5 minutes, then analysed by restriction enzyme digestion.

- Lanes 1- Undiluted S173C resolvase, 1 hour
- 2 - 2-fold dilution of S173C resolvase, 1 hour
- 3 - 4-fold dilution of S173C resolvase, 1 hour
- 4 - 8-fold dilution of S173C resolvase, 1 hour
- 5 - 16-fold dilution of S173C resolvase, 1 hour
- 6 - 32-fold dilution of S173C resolvase, 1 hour
- 7-12 as lanes 1-6 + *Bgl*II
- 13 - Undiluted Y176C resolvase, 24 hours
- 14 - 2-fold dilution of Y176C resolvase, 24 hours
- 15-16 as lanes 13-14 + *Bgl*II
- 17 - Undiluted K177C resolvase, 1 hour
- 18 - 2-fold dilution of K177C resolvase, 1 hour
- 19 - 4-fold dilution of K177C resolvase, 1 hour
- 20 - 8-fold dilution of K177C resolvase, 1 hour
- 21 - 16-fold dilution of K177C resolvase, 1 hour
- 22 - 32-fold dilution of K177C resolvase, 1 hour
- 23-28 as lanes 12-18 + *Bgl*II

resolvase concentration was low. The A171C, S173C and K177C proteins were able to bind, form a productive synapse and recombine the substrate DNA with similar efficiency to the wild-type protein. The engineered cysteine mutants at positions 171, 173 and 177 therefore do not affect the ability of the proteins to catalyse the recombination, although they may not bind as efficiently as wild-type resolvase.

The mutants R172C and Y176C show no resolution activity but display topoisomerase activity as observed in Figure 4.17 (see lanes 17-20) for R172C and Figure 4.18 (lanes 8 and 9) for Y176C. The ladder of bands, or topoisomers arose as a result of relaxation of negative supercoils in pMM5. This occurs when only one phosphodiester bond is broken, whereas recombination requires four strand breaks. Even though they were deficient at promoting recombination, Y176C and R172C retained the ability to promote cleavage and re-ligation, displayed by the observed topoisomerase activity.

A more important consideration is whether or not the introduction of the cysteine residues in the DNA-binding domain will affect binding. The above positive recombination results show that the presence of cysteine thiols at positions 171, 173 and 177, do not completely prevent binding. A171C, S173C and K177C still have the ability to bind to the DNA in the correct geometric arrangement and with enough efficiency to allow a productive synaptic complex to be formed. Once this synaptic structure is formed, the catalysis and the strand exchange reaction can occur resulting in a -2 catenane. The cysteines at 172 and 176 show topoisomerase activity, that similarly requires binding to all or part of *res*.

#### **4.12 Band shift analysis of mutants**

To determine the binding activities of the novel cysteine mutants of Tn3 resolvase, binding assays were carried out. It was important to know the extent to which the modified proteins bind to the DNA, in order to determine the effect the binding efficiency might have for crosslinking.

The binding assay (or gel retardation assay) displays the different mobilities of complexes that are formed between *res* and resolvase, as they move through polyacrylamide gels. Complexes having one or two protein monomers attached (i.e. one subsite of *res* fully occupied) run much faster than higher complexes, having the three subsites of *res* occupied by up to six protein monomers.

In-depth binding analysis of *res* subsite II (Blake *et al.*, 1995) indicated that Tn3 resolvase is monomeric in solution and binds to the subsites in a co-operative manner. A protein monomer is thought to bind to the left end of site II, and when a second monomer associates with this monomer-binding site complex, dimerisation is completed. Six complexes can be seen in binding assays of Tn3 resolvase with *res* (Blake *et al.*, 1995).

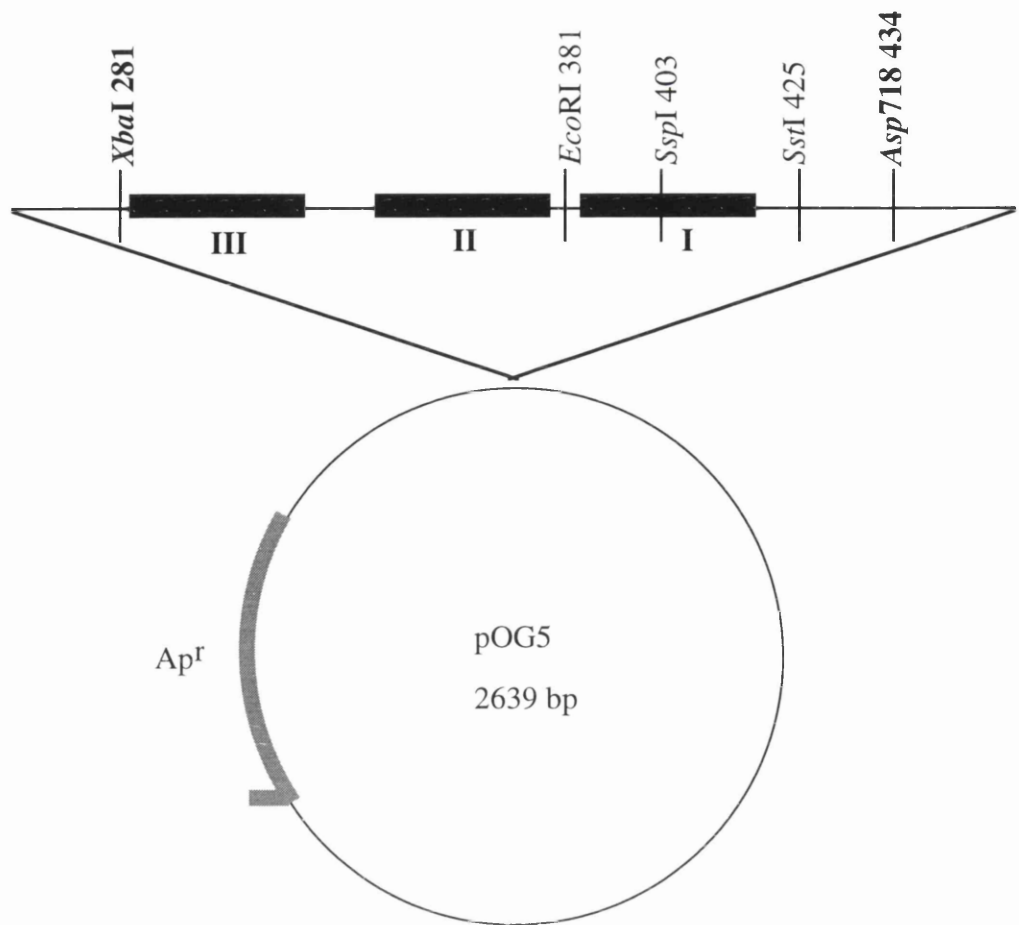
It is known from the co-crystal structure of  $\gamma\delta$  resolvase that a dimer is present at site I, and all binding of  $\gamma\delta$  resolvase to the other *res* subsites is by pre-formed dimers. Only three complexes, corresponding to each subsite of *res* occupied by dimers, were observed (Abdel-Meguid, 1984; Blake, 1993). The  $\gamma\delta$  resolvase exists in a dimeric conformation in solution, indicated by the binding profiles and by the results of crystallographic analysis of the N-terminal domains (Sanderson *et al.*, 1990) and by dimerisation studies (Liu *et al.*, 1993).

#### **4.13 Binding of resolvase cysteine mutants to *res* DNA**

The labelled *res* site was derived from the plasmid pOG5, which contains a single *res* site, illustrated in Figure 4.19. pOG5 was cut with *Xba*I then end-labelled (Section 3.18). The linear fragment was then cut with *Asp*718, releasing the labelled *res* site.

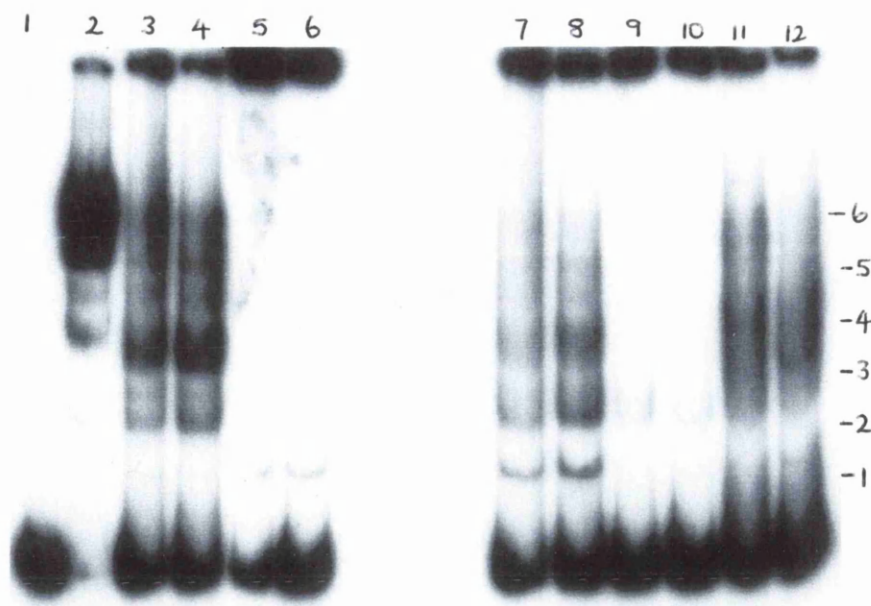
Figure 4.20 shows binding of the wild-type and all the mutant resolvases to *res*. Lane 1 contains labelled fragment only. Lane 2 contains the wild-type protein as a positive control. The wild-type control shows full occupancy of all six binding sites in *res*.





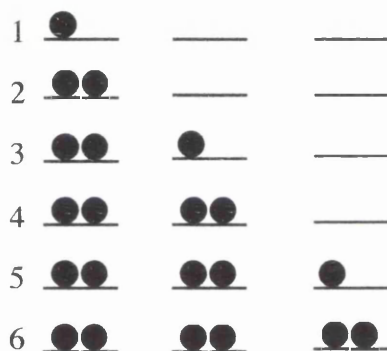
**Figure 4.19 Diagram of pOG5**

The plasmid shown provided the *res* fragment that was used in the binding assays. The fragment was cut from pOG5 by an *Xba*I-*Asp*718 restriction digest. Three base pairs between sites I and II have been mutated in pOG5 to create an *Eco*RI site (Blake *et al.*, 1995).



**Figure 4.20 Binding of A171C, R172C, S173C, Y176C and K177C resolvases to *res***

The numbers denote the following binding site \_\_\_\_\_ occupancy by resolvase ●



- Lanes 1 - No resolvase (binding buffer)  
 2 - 16-fold dilution of wild-type resolvase  
 3 - Undiluted A171C resolvase  
 4 - 2-fold dilution of A171C resolvase  
 5 - Undiluted R172C resolvase  
 6 - 2-fold dilution of R172C resolvase  
 7 - Undiluted S173C resolvase  
 8 - 2-fold dilution of S173C resolvase  
 9 - Undiluted Y176C resolvase  
 10 - 2-fold dilution of Y176C resolvase  
 11 - Undiluted K177C resolvase  
 12 - 2-fold dilution of K177C resolvase

In lanes 3 and 4 (A171C) there is a very intense band representing complex 2, indicating that the resolvase monomers bind co-operatively. There are also some higher complexes formed by the protein binding to more subsites of *res*, although weaker than complex 2. Approximately half of the labelled *res* DNA is bound in total. The strong band of complex 2 indicates that mutation at position 171 does not disrupt co-operative binding to *res*.

Lanes 5 and 6 show the binding of the R172C. Binding is very poor, with no complexes detected. This indicates that an interaction that is essential for binding under normal conditions has been disrupted by the introduction of a cysteine residue at this position. Some of the R172C-bound DNA is stuck in the wells. This may be due to aggregation following aberrant binding. The faint band could indicate binding of a monomer (as seen in lanes with 7 and 8, and 11 and 12 with different mobility to A171C and wild-type complexes).

Lanes 7 and 8 represent the complexes formed between S173C and *res*. The bands representing complex 1 and 2 appear to be almost equal in intensity. This may indicate loss of co-operative binding between the protein monomers. There were overall fewer complexes present than with A171C, indicating that this mutation is having a more inhibitory effect.

The Y176C binding profile is shown in lanes 9 and 10. At both concentrations there were very low levels of binding complexes on the gel. A weak band representing complex 2 indicated that a small amount of binding may be occurring.

Lane 11 and 12 represents the K177C mutant binding to *res*. At the highest concentration (lane 11) there are two main bands representing complex 2, and a slightly weaker higher band representing complex 4. The apparent widening of the bands and smearing of the complexes could be due to the dissociation of the non-covalent protein-DNA complexes during the electrophoresis, perhaps

because the mutation causes a weaker binding interaction than with wild-type protein.

#### **4.14 Binding of resolvase cysteine mutants to site 1 DNA**

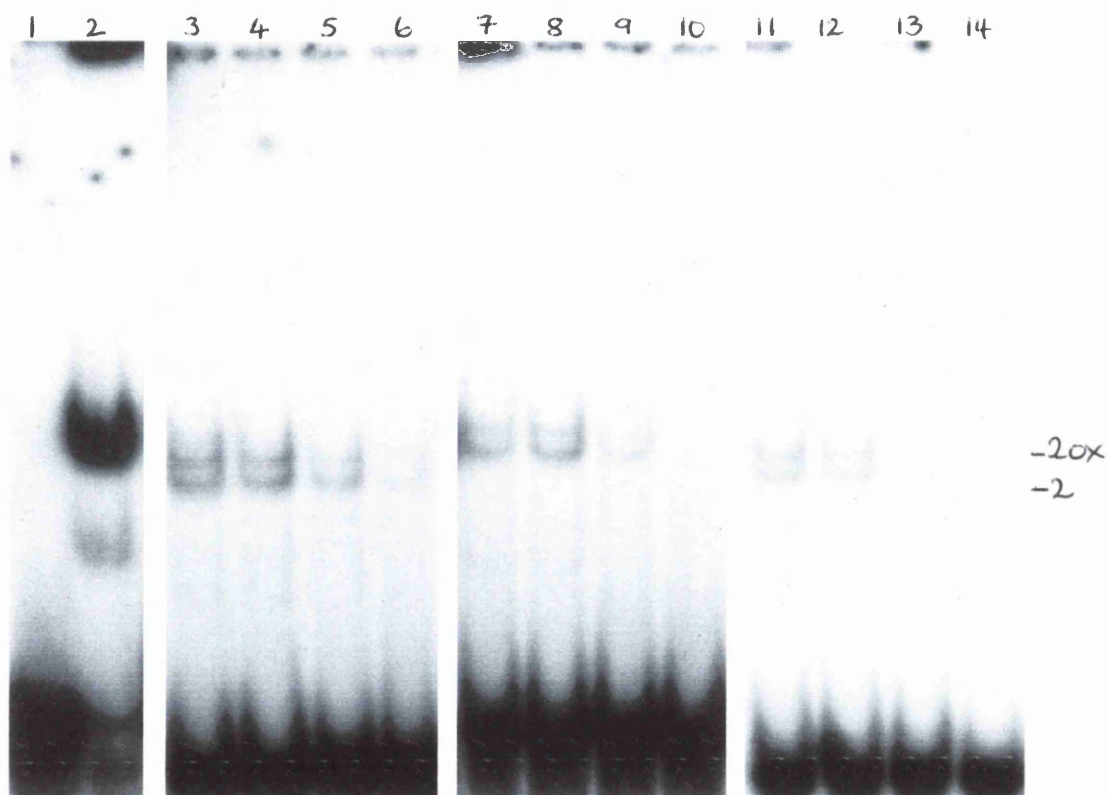
It was important to measure the binding of the cysteine mutants to site I DNA, as it was this site that was to be modified with the crosslinking group. An important pre-requisite to crosslinking was that there was sufficient binding of the proteins to *res* subsite I. Oligonucleotides (40 nt top strand and 48 nt bottom strand) forming wild-type subsite I were radiolabelled with  $^{32}\text{P}$  (Section 3.16).

Figure 4.21 shows binding of A171C, S173C and K177C to labelled site I. The wild-type control is shown in lane 2. The expected binding complex 1 was observed, corresponding to the two resolvase monomers bound at site 1. One of the bands represents the expected complex 2, while the other may be some oxidised form as illustrated in Figure 4.22 which may (or not) be more retarded than wild-type complex 2. It could be that the 'normal' complex 2 would run closer to the wild-type complex 2 and that the 'oxidised' complex 2 would be more retarded.

The R172C and Y176C resolvases form no detectable complexes with site I DNA (results not shown). This was expected, as no complexes were observed with binding reactions using *res* DNA.

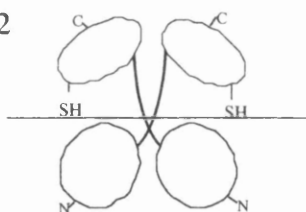
There is overall less binding to site I DNA compared to *res* DNA with the cysteine mutants. This could be due to the site I complexes being more unstable than complexes of the full *res* site, causing them to dissociate under the gel electrophoresis conditions.

The mutants A171C, S173C and K177C are able to bind to both *res* and site I DNA, whereas R172C and Y176C have greatly reduced binding to *res* DNA and are unable to bind to site I DNA sufficiently to see complexes under the assay

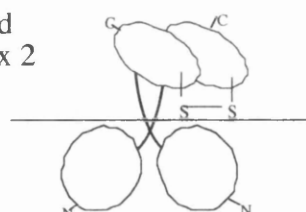


**Figure 4.21 Binding of A171C, S173C and K177C resolvases to site I**

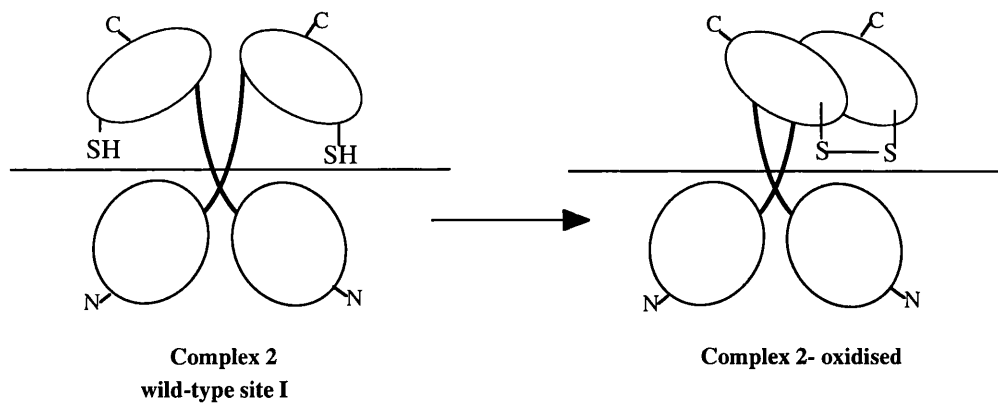
2 = complex 2



2ox = oxidised complex 2



- Lanes 1 - No resolvase (binding buffer)  
 2 - 16-fold dilution of wild-type resolvase  
 3 - Undiluted A171C resolvase  
 4 - 2-fold dilution of A171C resolvase  
 5 - 4-fold dilution of A171C resolvase  
 6 - 8-fold dilution of A171C resolvase  
 7 - Undiluted S173C resolvase  
 8 - 2-fold dilution of S173C resolvase  
 9 - 4-fold dilution of S173C resolvase  
 10 - 8-fold dilution of S173C resolvase  
 11 - Undiluted K177C resolvase  
 12 - 2-fold dilution of K177C resolvase  
 13 - 4-fold dilution of K177C resolvase  
 14 - 8-fold dilution of K177C resolvase



**Figure 4.22 Doublet band in binding gels**

Representation of the possible structure of the slower moving band present in site I binding gels. The oxidised species would be more retarded because of an altered shape of the DNA binding domains. N denotes the N-terminal domain of resolvase and C indicates the C-terminal DNA binding domain.

conditions used. However R172C and Y176C may still be in close enough proximity to the DNA, but the interaction is not strong enough to remain intact during the conditions used here. The fact that there are binding complexes detected with site I for A171C, S173C and K177C make them good candidates for crosslinking. Although R172C and Y176C show almost no binding under these assay conditions, they may still bind weakly to the DNA, whereby they could form covalent crosslinks between the proteins and the DNA if the two thiols in close proximity form disulfides.

#### 4.15 Discussion

A171C, S173C and K177C show both binding and recombination activity, whereas R172C and Y176C show little binding activity, and no recombination, with topoisomerase activity after extended reaction times. The topoisomerase activity of resolvase is proposed to occur by rotation of the left half of the cleaved site I, after all four half sites have been cleaved (Boocock *et al.*, 1995). It is more likely that the left half rotates, as the right half of site I is directly influenced by, and is thought to be structurally maintained by interaction of dimers bound to sites II and III.

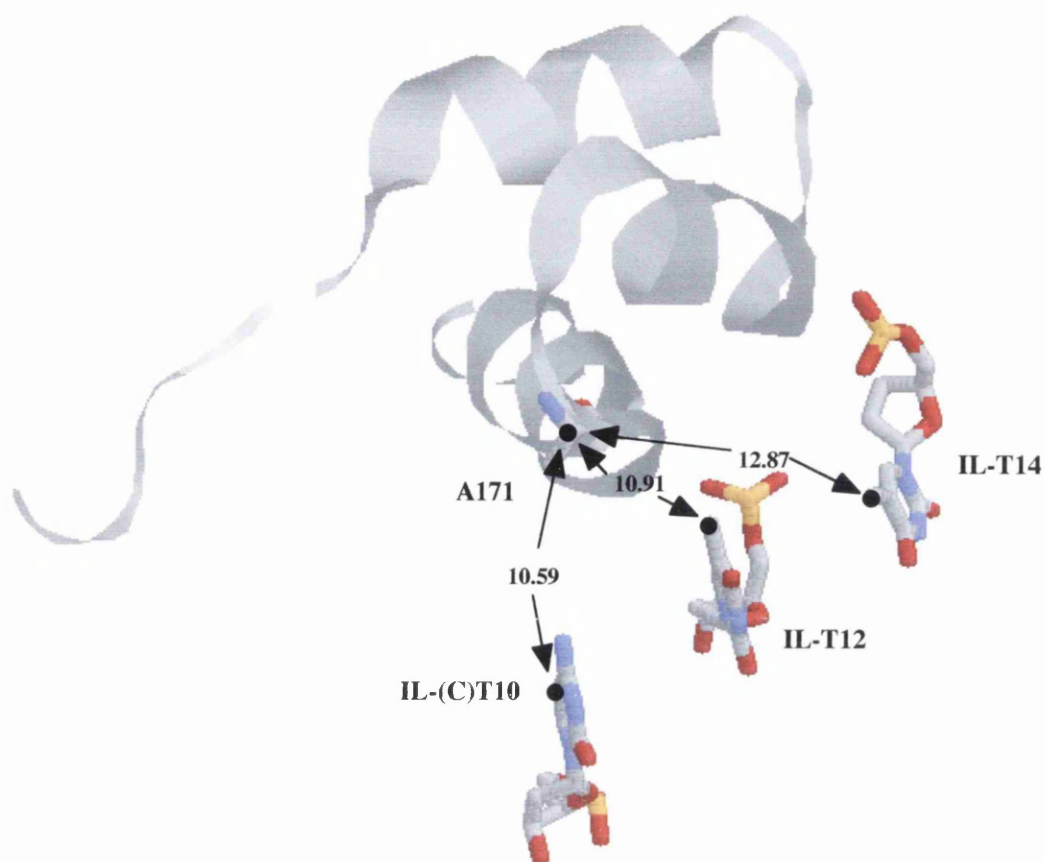
The results of the *in vitro* assay (Section 4.11) correlate well with the *in vivo* studies (Section 4.7). The proteins that were able to carry out recombination *in vivo*, could do so *in vitro*, and also formed detectable protein-DNA-binding complexes. Similarly, the proteins that failed to recombine *in vivo* formed no detectable protein-DNA complexes and showed no full recombination activity *in vitro*. Even when there is no observed recombination and/or binding activity, the R172C and Y176C mutants may still interact with the DNA sufficiently to be crosslinked.

Although A171 does not interact with the DNA in the crystal structure of the protein-DNA complex (Yang and Steitz, 1995), it is still theoretically close enough to the DNA for crosslinking to be carried out. It was predicted that

mutation of this residue should not have any great impact on binding and recombination in the A171C resolvase mutant. The *in vivo galK* assay for binding (Section 4.7) showed a positive result with the production of white colonies on MacConkey agar plates, showing that the A171C protein can catalyse recombination in cells. The results of the binding and recombination assays showed that the presence of the cysteine residue at the 171 position does not cause major unfavourable obstructions to the binding or recombination reaction. *In vitro* recombination by A171C is as efficient as wild-type resolvase *in vitro*. The binding gel of A171C to *res* (see Figure 4.20) shows strong bands corresponding to complexes 2 and 3, with a smeary band representing where the higher complexes would run. The binding profile is different to that of the wild-type protein, indicating that the mutation of this position caused changes that do not allow binding to occur naturally. This suggests that A171 in the protein is essential for a normal binding interaction to occur, so it may in fact interact with the DNA, perhaps by indirect methods. The distance between the  $\beta$ -carbon of A171 and the C-5 of thymidine is 12.87 Å for IL-T14, 10.49 Å for IL-T12 and 10.55 Å for IL-C(T)10 (Figure 4.23).

Mutation of R172 to cysteine resulted in a dramatic decrease in binding to the DNA, although a very low level of complex 2 was observed with a labelled *res* site. Although it is inactive *in vivo* and shows little binding activity in the gel assay, the topoisomerase activity of R172C indicates that the mutant may still be a candidate for crosslinking. The fact that some binding and single-stranded cleavage of the DNA is occurring as observed in Figure 4.18, means that the formation of a disulfide between the protein thiol and the modified DNA is feasible. The modification R→C replaces a charged long chain residue with the cysteine methyl thiol. In the co-crystal structure, the arginine at position 172 interacts with IR-T14 and IR-G13 in the consensus TGT binding motif. The arginine guanidinium nitrogen atoms interact with the O6 and N7 of IL-G13, and the aliphatic (CH<sub>2</sub>)<sub>3</sub> portion of the arginine makes hydrophobic interactions with the methyl group of IL-T14. One would expect a reduction in activity as all of the





**Figure 4.23 Interaction of A171 with modified bases**

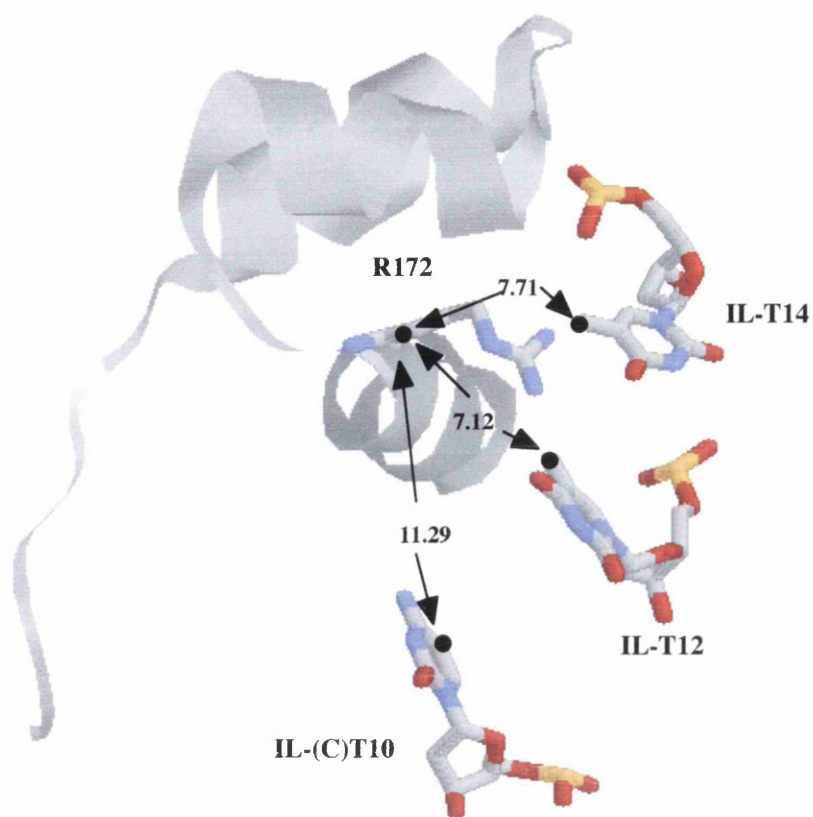
The three positions of the DNA that are modified relative to the A171 residue. The black dots indicate the positions that will have thiol groups. The distances between the thiols are numbered. Atoms in CPK (Corey-Pauling-Koltun) colours. The data are derived from Rasmol.

wild-type interactions are abolished by the cysteine mutation. The calculated distance for crosslinking is 7.71 Å for IL-T14, 7.12 Å for T12 and 11.29 Å for C(T)10 (Figure 4.24).

Like A171C, the S173C mutant displayed *in vivo* activity. The change S→C is quite conservative, the electronegative oxygen atom of the natural serine replaced with a sulfur atom. The electronegativity of the atom at this position is maintained. However, replacement of oxygen with sulfur will cause a size change due to the larger atomic radius of sulfur. From analysis of the co-crystal structure it was observed that the serine at position 173 differs in its interaction with each of the two DNA-binding domains in the complex. The hydrogen of the serine hydroxyl group is involved in hydrogen bonding to the O6 of IR-G10 and IR-G11 and IR-G10' in one subunit only (Figure 4.25).

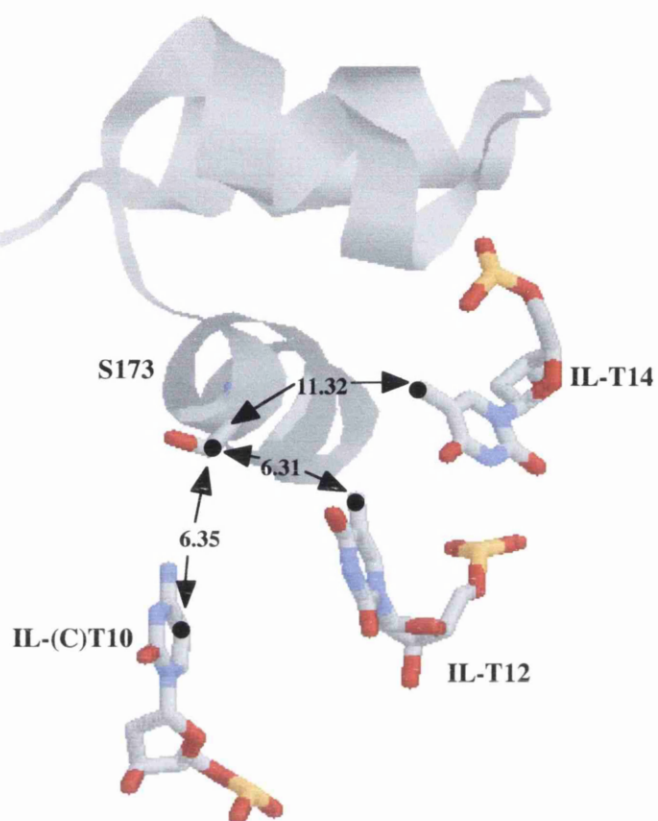
Replacing the hydroxyl group with a cysteine thiol should not necessarily disrupt binding interactions, as the cysteine thiol hydrogen could still participate in hydrogen bonding to the IL-G10 or IL-G11. The crosslinking potential at this position is high, since binding is not prevented by the presence of the sulfur atom. The β-carbon atom of S173 is 11.32 Å away from IL-T14, 6.31 Å away from IL-T12 and 6.35 Å away from IL-C(T)10 (Figure 4.25).

The tyrosine at position 176 is the closest to the DNA of the five residues chosen to be mutated. This tyrosine forms hydrophobic interactions between the β-carbon atom and the methyl group of IL-T12. The mutation Y→C introduces a sulfur atom and a dramatic size change; the steric bulk induced by the phenol ring in the natural base is lost. The change in polarity and differences in conformation of the Y176C mutant lead to disruption of all the wild-type interactions. The β-carbon of Y176 is 9.81 Å away from IL-T14, 5.14 Å away from IL-T12 and 9.94 Å away from IL-C(T)10 (Figure 4.26).



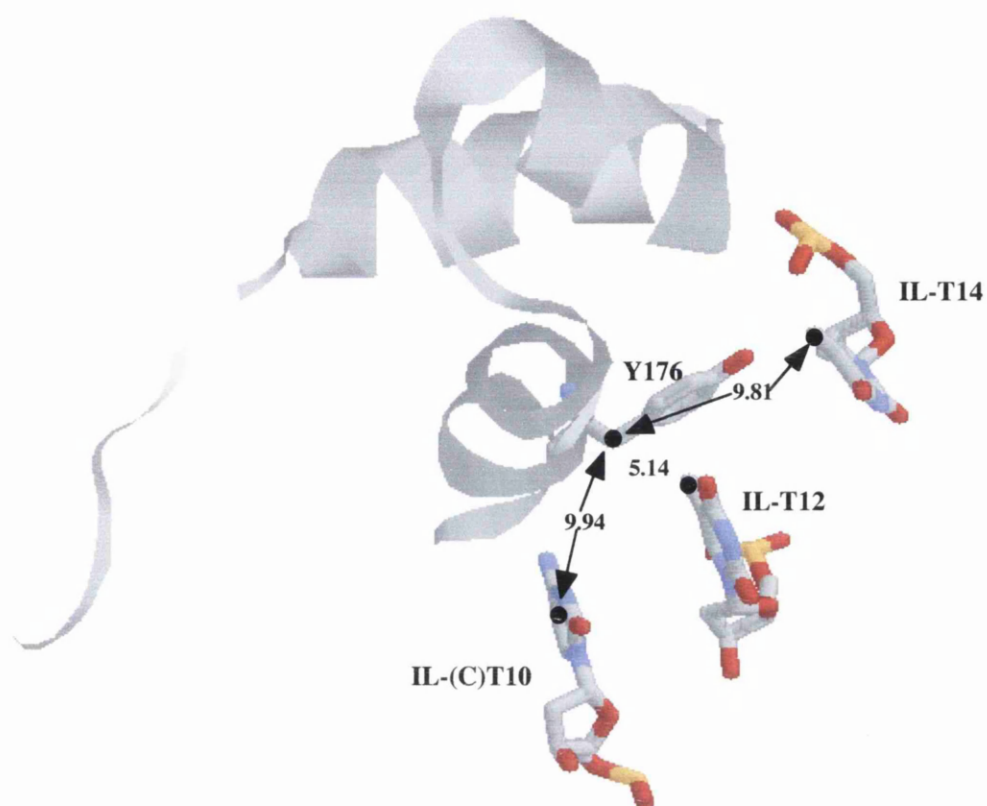
**Figure 4.24 Interaction of R172 with modified bases**

The three positions of the DNA that are modified relative to the R172 residue. The black dots indicate the positions that will have thiol groups. The distances between the thiols are numbered. Atoms in CPK (Corey-Pauling-Koltun) colours.



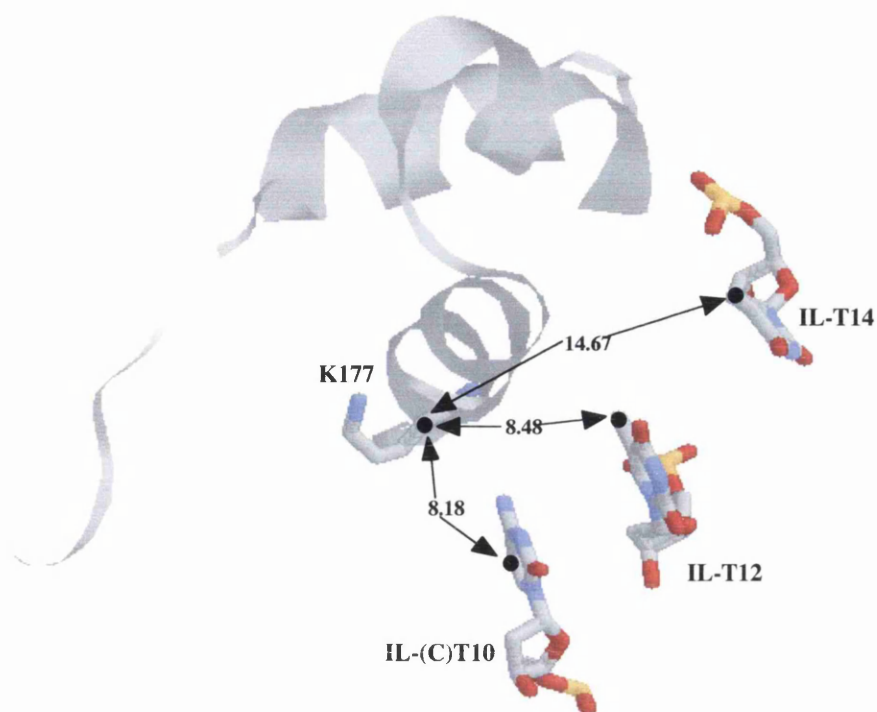
**Figure 4.25 Interaction of S173 with modified bases**

The three positions of the DNA that are modified relative to the S173 residue. The black dots indicate the positions that will have thiol groups. The distances between the thiols are numbered. Atoms in CPK (Corey-Pauling-Koltun) colours.



**Figure 4.26 Interaction of Y176 with modified bases**

The three positions of the DNA that are modified relative to the Y176 residue. The black dots indicate the positions that will have thiol groups. The distances between the thiols are numbered. Atoms in CPK (Corey-Pauling-Koltun) colours.



**Figure 4.27 Interaction of K177 with modified bases**

The three positions of the DNA that are modified relative to the K177 residue. The black dots indicate the positions that will have thiol groups. The distances between the thiols are numbered. Atoms in CPK (Corey-Pauling-Koltun) colours.

The lysine residue at position 177 forms hydrophobic interactions with the methyl group of IR-T8, and a salt-bridge between the phosphate backbone of the DNA and its amine group. Like A171C and S173C, the K177C mutant showed recombinational activity *in vivo* and *in vitro*, and formed protein-DNA complexes with both *res* and site I DNA. The mutation K→C replaces a four carbon chain, carrying a charged amino group, with an electronegative mercaptomethyl group. Therefore in the K177C mutant the hydrophobic interaction is lost, as is the formation of the salt bridge. The binding gel of K177C showed very smeary complexes indicating that the interaction is somewhat unstable. The changes in polarity could mean that the binding interaction in the mutant protein is much weaker than the wild-type protein, resulting in aberrant binding complexes. The  $\beta$ -carbon of K177 is 14.67 Å away from IL-T14, 8.48 Å away from IL-T12 and 8.18 Å away from IL-C(T)10 (Figure 4.27).

#### 4.16 Conclusions

The novel Tn3 cysteine mutants purified and characterised in this chapter are all good candidates for crosslinking. Three out of the five, A171C, S173C and K177C are still able to bind well to *res* DNA, which is essential for crosslinking. R172C and Y176C do not bind strongly, but may be come close physically close enough during the reaction to be crosslinked.

## **Chapter 5: Crosslinking using modified oligonucleotides and cysteine mutants**



## 5.1 Uses of modified oligonucleotides

Oligonucleotides can be modified at a number of different sites. These modifications can be introduced at the 5' and 3' termini, the internucleotide linkage, the sugar group or the bases (reviewed in Goodchild, 1990; Verma and Eckstein, 1998). The oligonucleotides used here were modified with a thiol-containing base analogue protected as the phosphoramidite (reviewed in Beaucage and Iyer, 1993).

## 5.2 Oligonucleotides containing the DNP group, T1-T4 and B1-B5

5'-*O*-(Dimethoxytrityl)-5-[*S*-(2,4-dinitrophenyl)-thio]-2'-deoxyuridine, 3'-*O*-(2-*N,N'*-cyano-ethyl diisopropylphosphoramidite **SH 17** was incorporated into oligonucleotides with sequences corresponding to *res* subsite I (illustrated in Figure 5.1). The phosphoramidite obtained from phosphoramidite method B (Chapter 2) was used in all the oligonucleotide syntheses, as it was the best quality product. The top strand 40-mer and the bottom strand 48-mer were synthesised on an Applied Biosystems 392 DNA synthesiser (Section 3.38). Following the removal of the crude oligonucleotides from the CPG support, the nucleotides were deprotected, purified and analysed for the presence of the 5-dinitrophenyl (DNP) group. The positions T1, T3, T4, B1, B3 and B5 were chosen for modification because they are within reasonable distances of residues in the resolvase recognition helix, that were mutated to cysteine (Chapter 4). The oligonucleotides T12 and B13 were synthesised as controls.

## 5.3 UV Spectrometry of the DNP modified oligonucleotides

Oligonucleotides that have no modifications show two main UV absorptions at 215 and 260 nm. Initial measurements of the absorbance of the oligonucleotides containing **SH 17** showed little absorbance above 300 nm, in the region where the DNP group would absorb **SH P12** ( $\lambda_{\text{max}} = 315 \text{ nm}$ ,  $\epsilon = 24,500 \text{ l mol}^{-1} \text{ cm}^{-1}$ ).

Taking into account that only one base in 40 (for top strand modifications) carries the DNP group, the predicted UV spectrum of a 40-mer oligonucleotide

(1)

Sst I CACTAGTC**TGTT**CGAAATATTATAAATTATCAGACATAGGEco RI  
TCGAGTGATCAGACAAGCTTTATAATA**TT**TAATAG**TCTGT**ATCCTTAA

(2)

Top Strands

CACTAGTC**TGTT**CGAAATATTATAAATTATCAGACATAGG = **T1**

CACTAGTCTG**TT**CGAAATATTATAAATTATCAGACATAGG = **T3**

CACTAGTCTGT**TC**GAAATATTATAAATTATCAGACATAGG = **T4**

CACTAGTCTGTT**CGAA**ATA**TT**ATAAATTATCAGACATAGG = **T12**

Bottom Strands

AATTCCTA**TGT**CTGATAATTTATAATATTT**CGA**ACAGACTAGTGAGCT = **B1**

AATTCCTATG**TCT**GATAATTTATAATATTT**CGA**ACAGACTAGTGAGCT = **B3**

AATTCCTATGT**CTG**AATAATTTATAATATTT**CGA**ACAGACTAGTGAGCT = **B5**

AATTCCTATGTCTGATAATTT**TATA**AATATTT**CGA**ACAGACTAGTGAGCT = **B13**

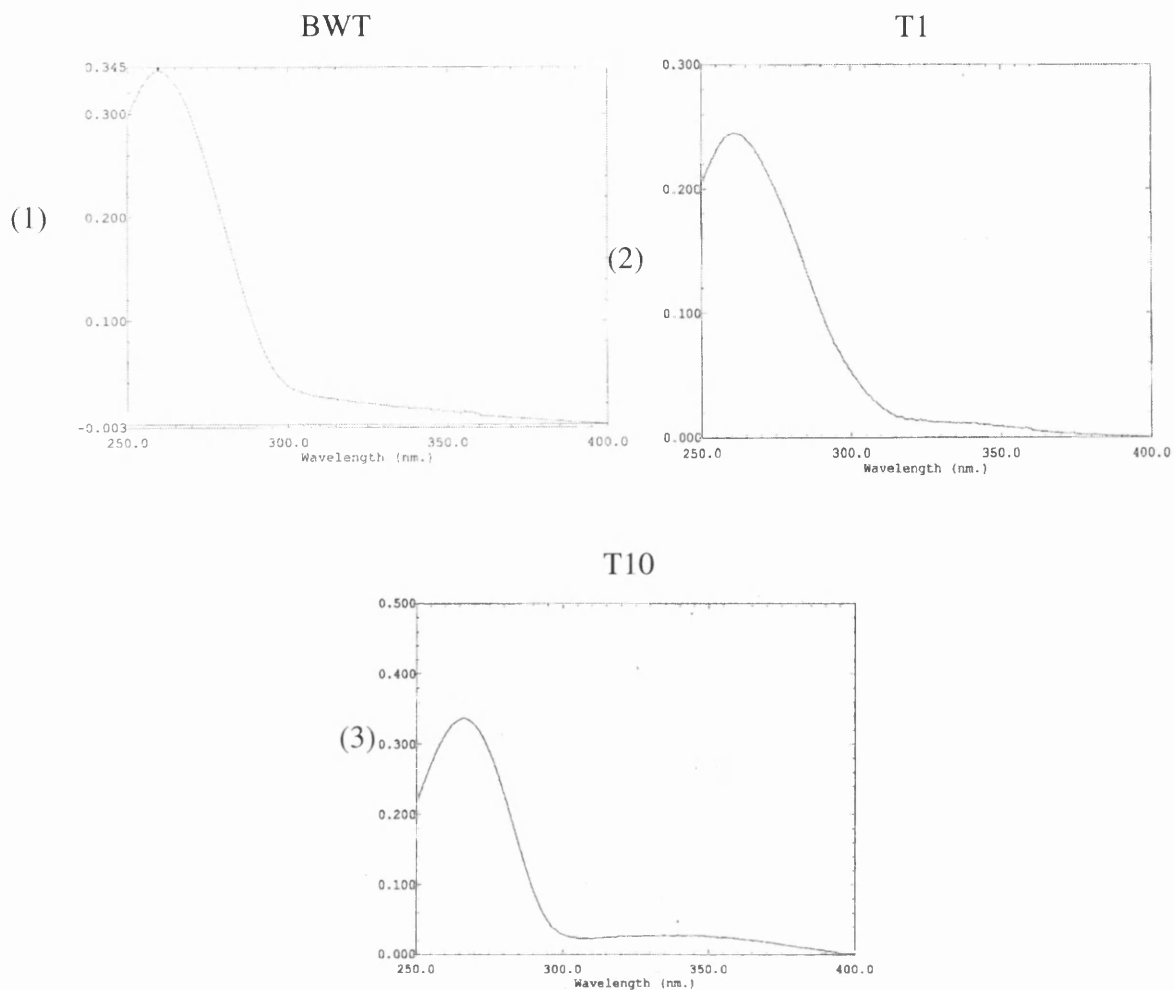
(3)

TT**TTTTTTTT** = **T10**

**Figure 5.1 Site I oligonucleotides**

(1) Illustration of the synthetic site I, flanked by *Sst*I and *Eco*RI sites. The thymidines in bold were substituted individually by **SH 17** (see (1)).

(2) The individual top and bottom strands, the position of the **SH 17** analogue is highlighted in bold. (3) Shows the control thymidine only **T10**.



**Figure 5.2 UV spectra of DNP modified oligonucleotides**

(1) UV absorbance of a wild-type *EcoRI-SstI* bottom strand wild-type oligonucleotide. (2) The absorbance of a top strand 40-mer **T1**, with the first T in TGT containing **SH 17** (3) spectrum of the **T10** modified oligonucleotide.

containing one DNP-modified thymidine should show a large absorbance at 260 nm, and a peak corresponding to approximately 5% of the intensity of the 260 nm peak at 330 nm. It was found, however, that there was no detectable difference between the UV spectra of purified wild-type oligonucleotides and oligonucleotides containing the DNP modification (Figure 5.2, compare spectra 1. and 2.). Indeed, even UV measurements using the ammonia stock solutions of the modified oligonucleotides, that had been cleaved from the columns and taken directly from storage at -70 °C, showed no obvious peaks above 300 nm. The absorbance of the DNP group was probably not detected because of the high levels of background absorbance from other chromophores.

In order to establish the presence of the DNP group in the modified oligonucleotides using UV spectrometry, a smaller oligonucleotide was synthesised. A 10-mer oligonucleotide containing 9 thymines, **SHT10**, one of which was modified with the phosphoramidite **SH 17** was synthesised. Thymidines have no exocyclic amide protecting groups, therefore the long ammonia deprotection step is not required. Therefore, the chances that the DNP group will be lost (by hydrolysis or ammoniolysis) during the ammonia deprotection step are reduced. By using a shorter modified oligonucleotide in the UV measurements, the ratio of absorbances at 330 nm and 260 nm should be increased. The oligonucleotide **SHT10** should show an absorbance above 300 nm.

The UV spectra of wild-type and two DNP modified oligonucleotides are shown in Figure 5.2. An absorbance at 330 nm of **SHT10** was observed, corresponding to the DNP moiety. The extinction co-efficient ( $\epsilon$ ) of the DNP group is approximately 20,000, whereas for the natural bases a Figure of 10,000 is taken as the average value for the natural nucleotides. The signal for the DNP group should mirror that ratio and be approximately 20% of the large 260 nm peak. The absorbance of the 330 nm peak was actually 10% that of the 260 nm absorbance. This apparent reduction in absorbance at 330 nm could be due to the presence of

some oligonucleotides that do not contain the DNP group, i.e. some ammoniolysis or hydrolysis could have occurred during purification. Another explanation is that the value of  $\epsilon$  is reduced when the DNP group is in the oligonucleotide.

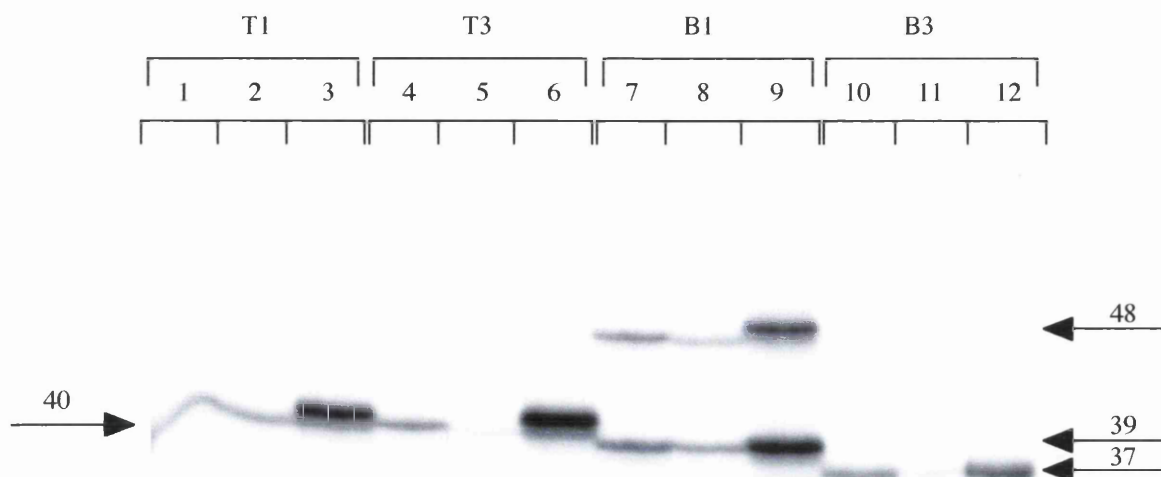
#### **5.4 Deprotection using $\beta$ -mercaptoethanol and alkylation using IAF**

The DNP group is removable by treatment with the reducing agent  $\beta$ -mercaptoethanol (Hanna and Meyer, 1996). Following the deprotection step, an alkylation reaction was carried out in order to confirm the presence of the free thiol group in the oligonucleotides. The oligonucleotides were deprotected in the first instance using 1.4 M  $\beta$ -mercaptoethanol (Section 3.40), then alkylated using 4(5)-(iodoacetamido)-fluorescein (Section 3.41). The products were separated on a sequencing gel (Section 3.24.4).

An autoradiograph of the gel is shown in Figure 5.3. Four different modified oligonucleotides were used, modified with phosphoramidite **SH 17** at different positions within the oligonucleotides. T1 and T3 are top strand oligonucleotides with the first and third thymidine of the TGT motif having a thiol-containing modification. B1 and B3 have the ninth and eleventh base of the bottom strand oligonucleotide bearing the DNP modification as illustrated in Figure 5.1.

For each oligonucleotide, samples before deprotection, after deprotection and after alkylation were run on the gel. The lanes containing oligonucleotides treated with  $\beta$ -mercaptoethanol will have lost the DNP group, and should have a mass that is 250 units less (the size of DNP). The lanes that have been treated with  $\beta$ -mercaptoethanol and subsequently alkylated using IAF should have gained a mass of 300 (molecular weight of IAF group).

Figure 5.3 (lanes 1-3) represent the deprotection/alkylation of oligonucleotide T1. In lane I the oligonucleotide has run quite unevenly, but a change in gel mobility is apparent between lanes 2 and 3. The band in lane 3 is more retarded than that



**Figure 5.3** Autoradiograph of deprotection and alkylation of T1, T3, B1 and B3

The third track in each set is taken up in a smaller volume (10  $\mu$ l) than the initial starting volume of the DNP-protected oligonucleotides (20  $\mu$ l). The sequences of T1, T3, B1 and B3 are shown in Figure 5.1. The arrows indicate the positions where one would expect the differently sized oligonucleotides to run. A reference wild type 40 mer was labelled and run at the edge of the gel (not shown) as a reference point for the size determination of the truncated oligonucleotides.

- Lane 1 - Untreated T1
- 2 - T1 treated with  $\beta$ -mercaptoethanol
- 3 - as lane 2, treated with IAF
- 4 - Untreated T3
- 5 - T3 treated with  $\beta$ -mercaptoethanol
- 6 - as lane 5, treated with IAF
- 7 - Untreated B1
- 8 - B1 treated with  $\beta$ -mercaptoethanol
- 9 - as lane 8, treated with IAF
- 10 - Untreated B3
- 11 - B3 treated with  $\beta$ -mercaptoethanol
- 12 - as lane 11, treated with IAF

of lane 2, indicating that the oligonucleotide has been alkylated. There is also a faster moving band in lane 3 with the same mobility as the thiol-containing oligonucleotide in lane 2, indicating that not all the free thiol was alkylated.

From Figure 5.3 (lanes 4-6) the observed changes in mobility are similar to the results obtained for T1. There is a band in lane 5 faster moving than that in lane 4, indicating that the oligonucleotide has lost the DNP group. Again the band in lane 6 is more retarded than lane 4, indicating that alkylation has occurred, and there is also a faster moving band, indicating that the thiol in T3 has not been completely alkylated.

The experiment with oligonucleotide B1 (Figure 5.3; lanes 7-9) showed that there are two species of oligonucleotide present. The top band is a 48-mer and corresponds to the full length oligonucleotide, where a change in size is observed on deprotection and subsequent alkylation, more noticeable between lanes 8 and 9 than in lanes 7 and 8. The lower set of bands are a 39-mer, having nine bases less than the full length oligonucleotide. This oligonucleotide is a result of failure of addition of the modified nucleotide during the DNA synthesis cycle. The modification in B1 is at the ninth position from the end of the oligonucleotide (Figure 5.1). Therefore, oligonucleotides in the lower bands in lanes 7-9 show no differences in size, because they have no DNP group present.

Similarly B3 has no full length 48-mer oligonucleotide; only a 37-mer is present. This result indicates that the DNA polymerisation reaction failed completely at the modified base. Oligonucleotides B1 and B3 were resynthesised for use in crosslinking studies.

In all lanes there is more labelled DNA in the third track, compared to the first two tracks. This is because after the alkylation step the DNA was taken up in a smaller volume of buffer than it was originally dissolved in, i.e. a 20  $\mu$ l initial

solution of DNP-protected oligonucleotide was taken up in a final volume of 10  $\mu\text{L}$ .

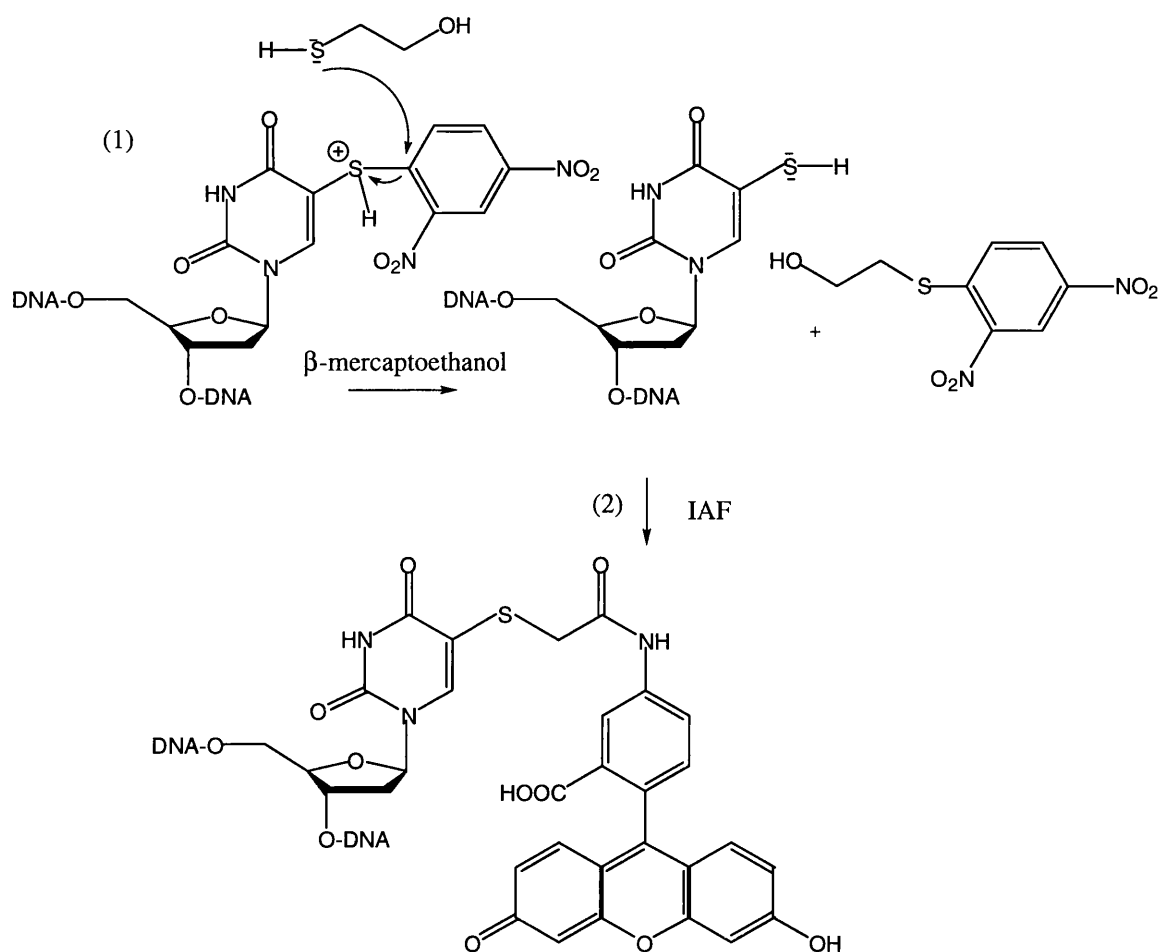
The chemical steps in the  $\beta$ -mercaptoethanol deprotection and the subsequent thiol alkylation are shown in Figure 5.4. A simple  $\text{S}_{\text{N}}2$  displacement of the iodine of IAF by the thiol results in the alkylation of the pyrimidine thiol. The mechanism for the  $\beta$ -mercaptoethanol deprotection is shown in Figure 5.4.

The results of the deprotection/alkylation experiment confirm the presence of the thiol in the oligonucleotides, indicated by the observed changes in mobility. The partial and complete failure of the DNA synthesis of the DNP-containing oligonucleotides could be due to chemical changes in **SH 17**. Extensive characterisation of the phosphoramidite **SH 17** made by method B (Chapter 2) indicated that all the protecting groups were present; this was confirmed by  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{32}\text{P}$  NMR, and accurate mass. The 4,4'-dimethoxytrityl (DMT) group on the 5'-hydroxyl is very acid labile and removed by treatment with trichloroacetic acid prior to the coupling step with the incoming phosphoramidite monomer. However, there was no evidence that this group was not present in **SH 17** (Chapter 2). However if the DMT group had been lost this would result in failure of the synthesis at the coupling step.

A possible problem with the 3'-hydroxyl phosphoramidite protecting group, is that some oxidation of the phosphorus may have occurred, as it is very sensitive to air oxidation causing  $\text{P}^{\text{III}}\text{-P}^{\text{V}}$ . However, great care was taken to ensure that **SH 17** was protected from atmospheric oxygen. Even if some of **SH 17** contained  $\text{P}^{\text{V}}$ , this would not prevent the coupling step as  $\text{P}^{\text{V}}$  is unreactive, but would probably cause a lower yield of the DNP-protected oligonucleotides.

A more important consideration is whether the integrity of the DNP protecting group in the oligonucleotides was compromised. In DNP-modified oligonucleotides that had been deprotected in ammonia for longer times, there was an observed colour change from yellow to clear, indicating that the DNP





**Figure 5.4 Chemistry of deprotection and alkylation of 5-mercaptouridine**

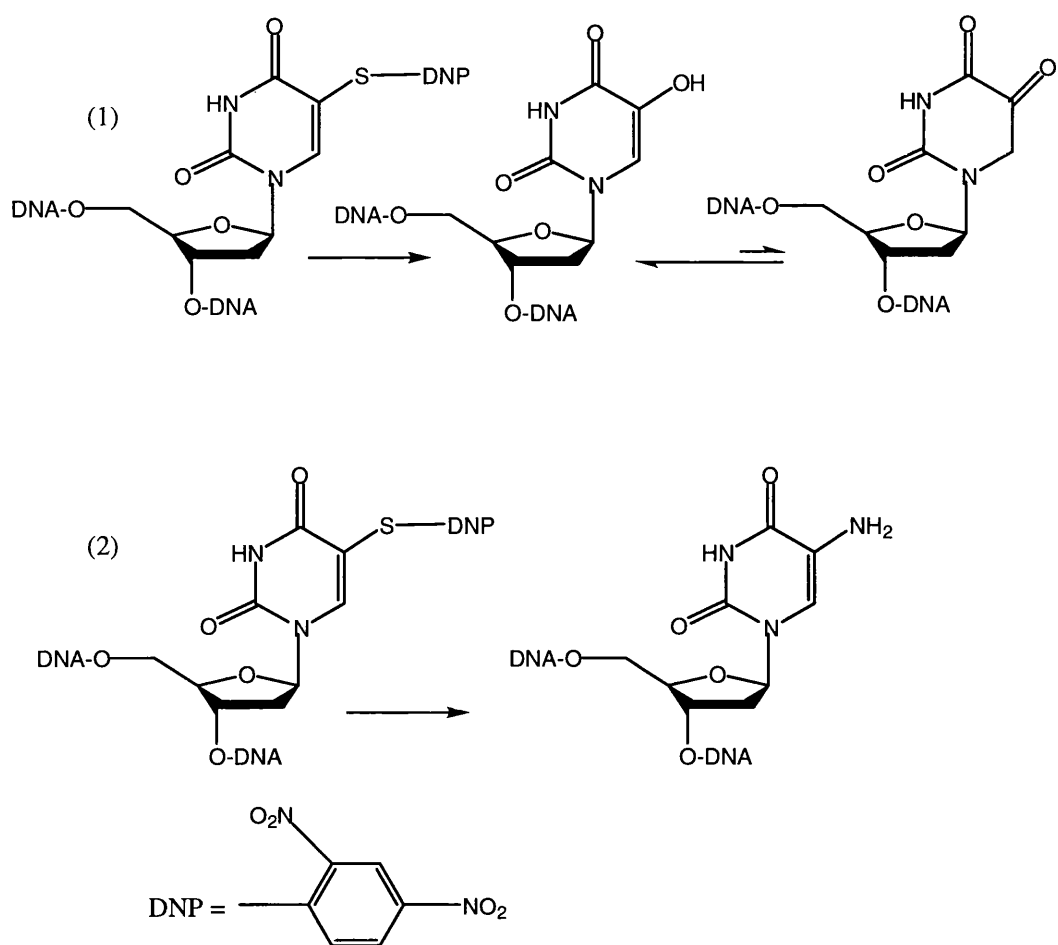
(1) The mechanism of removal of DNP, generating the free thiol. Step (2) is alkylation using 4(5)-(iodoacetamido)-fluorescein (IAF).

group was no longer present. The presence of thiol in the oligonucleotides was tested using Ellmans reagent [5, 5'-dithiobis(2-nitrobenzoic acid)]. This reagent should form a mixed disulfide with any thiol present and cause an intense yellow colour due to the liberation of 1 mole of nitrothiobenzoate per mole of thiol group present. The results of this test were negative (results not shown). It was also possible the DNP group may have formed *bis*(2,4-dinitrophenyl)disulfide, which is colourless. Addition of DTT should reduce this and form 2,4-dinitrobenzene thiolate ion that would be yellow. Again the test failed. It is more likely that hydrolysis or ammoniolysis of the DNP group in the modified oligonucleotides occurred, yielding by-products (Figure 5.5).

### **5.5 Binding of the modified oligonucleotides to the cysteine mutant proteins**

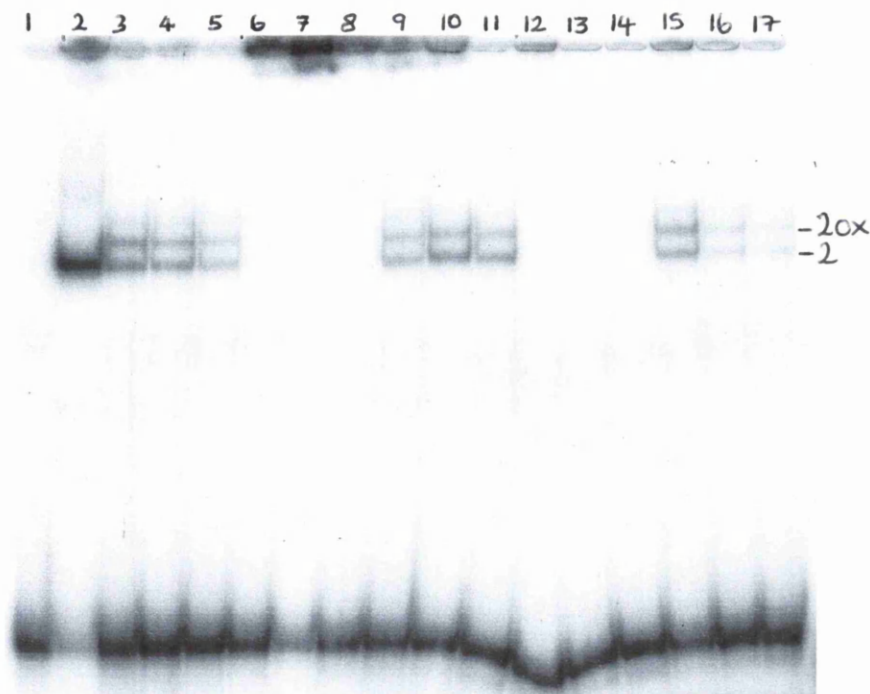
The oligonucleotides containing the DNP modification were end-labelled (Section 3.16), then annealed to the complementary wild-type strand (Section 3.20). It was important to show that the modified oligonucleotides could bind to the cysteine mutants, as it is essential for crosslinking. Binding analysis was performed with different concentrations of DTT in the binding mix, ranging from 1 mM to 5 mM. The results of binding at 5 mM are shown in Figures 5.6 and 5.7. The binding assay carried out in 2 mM DTT showed very weak binding with S173C; no complexes were observed for the other cysteine mutants (data not shown).

In all lanes in Figure 5.6, there is a band representing the unbound oligonucleotides. One of these bands (band 1) is double stranded oligonucleotides containing one thiol group. The other band, present in Figure 5.7 (band 2) is single-stranded oligonucleotide, due to not enough cold oligonucleotide being added to the annealing reaction, or the double-stranded oligonucleotides may have denatured prior to the gel being run. An oxidised dimer of oligonucleotides, less retarded than the oxidised oligonucleotides, was observed in all the crosslinking gels (Figure 5.10-5.16).



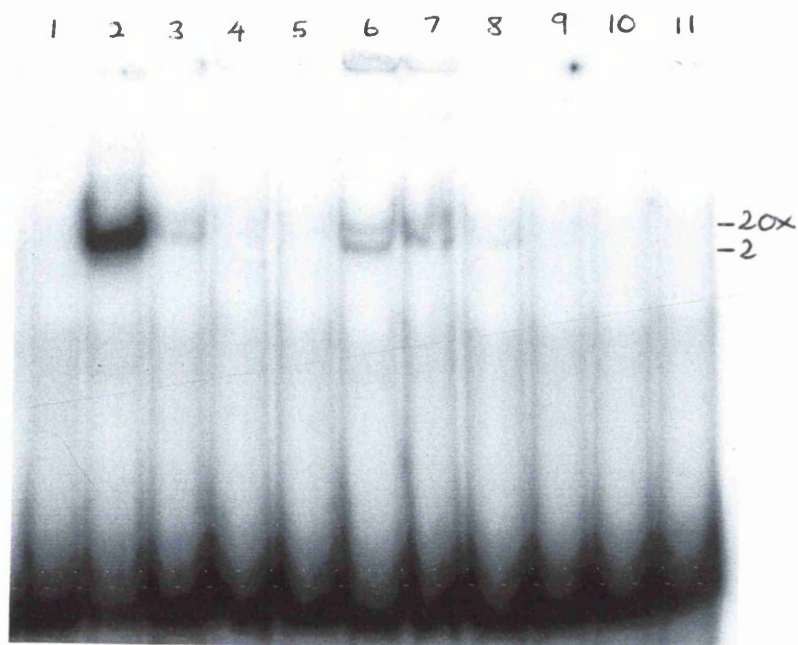
**Figure 5.5 Hydrolysis /ammoniolysis by-products of DNP-containing oligonucleotides**

Illustration of the possible by-products generated from hydrolysis of the DNP protected oligonucleotides. (1) shows hydrolysis resulting in the diketone, whereas (2) shows the 5-aminouracil by-products.



**Figure 5.6 Binding of double stranded oligonucleotide with modification at position B1 to A171C, R172C, S173C, Y176C and K177C (5 mM DTT)**

- Lanes
- 1 - No resolvase (binding buffer)
  - 2 - 16-fold dilution of wild-type resolvase
  - 3 - Undiluted A171C resolvase
  - 4 - 2-fold dilution of A171C resolvase
  - 5 - 4-fold dilution of A171C resolvase
  - 6 - Undiluted R172C resolvase
  - 7 - 2-fold dilution of R172C resolvase
  - 8 - 4-fold dilution of R172C resolvase
  - 9 - Undiluted S173C resolvase
  - 10 - 2-fold dilution of S173C resolvase
  - 11 - 4-fold dilution of S173C resolvase
  - 12 - Undiluted Y176C resolvase
  - 13 - 2-fold dilution of Y176C resolvase
  - 14 - 4-fold dilution of Y176C resolvase
  - 15 - Undiluted K177C resolvase
  - 16 - 2-fold dilution of K177C resolvase
  - 17 - 4-fold dilution of K177C resolvase



**Figure 5.7 Binding of double stranded oligonucleotide with modification at position B3 to A171C, S173C and K177C to B3 (5 mM DTT)**

- Lane 1 - No resolvase (binding buffer)
- 2 - 16-fold dilution of wild-type resolvase
- 3 - Undiluted A171C resolvase
- 4 - 2-fold dilution of A171C resolvase
- 5 - 4-fold dilution of A171C resolvase
- 6 - Undiluted S173C resolvase
- 7 - 2-fold dilution of S173C resolvase
- 8 - 4-fold dilution of S173C resolvase
- 9 - Undiluted K177C resolvase
- 10 - 2-fold dilution of K177C resolvase
- 11 - 4-fold dilution of K177C resolvase

Increasing the DTT concentration to 5 mM increased reduction of any oxidised forms of the cysteine mutants that may have been present in the samples, that may have impeded binding. Thus the cysteine monomers containing free thiol groups should give more binding complexes.

The binding assays resulted in complexes for A171C, S173C and K177C (Figure 5.6; lanes 3-5, 9-11 and 15-17). In general there were less binding complexes observed for the proteins, compared to wild-type resolvase (Figure 5.6; lane 2). A band corresponding to complex 2 was observed, although an additional retarded band was observed with all the mutants, but not with wild-type resolvase. This band was also observed in binding assays with wild-type site I oligonucleotides (Section 4.14; Figure 4.21).

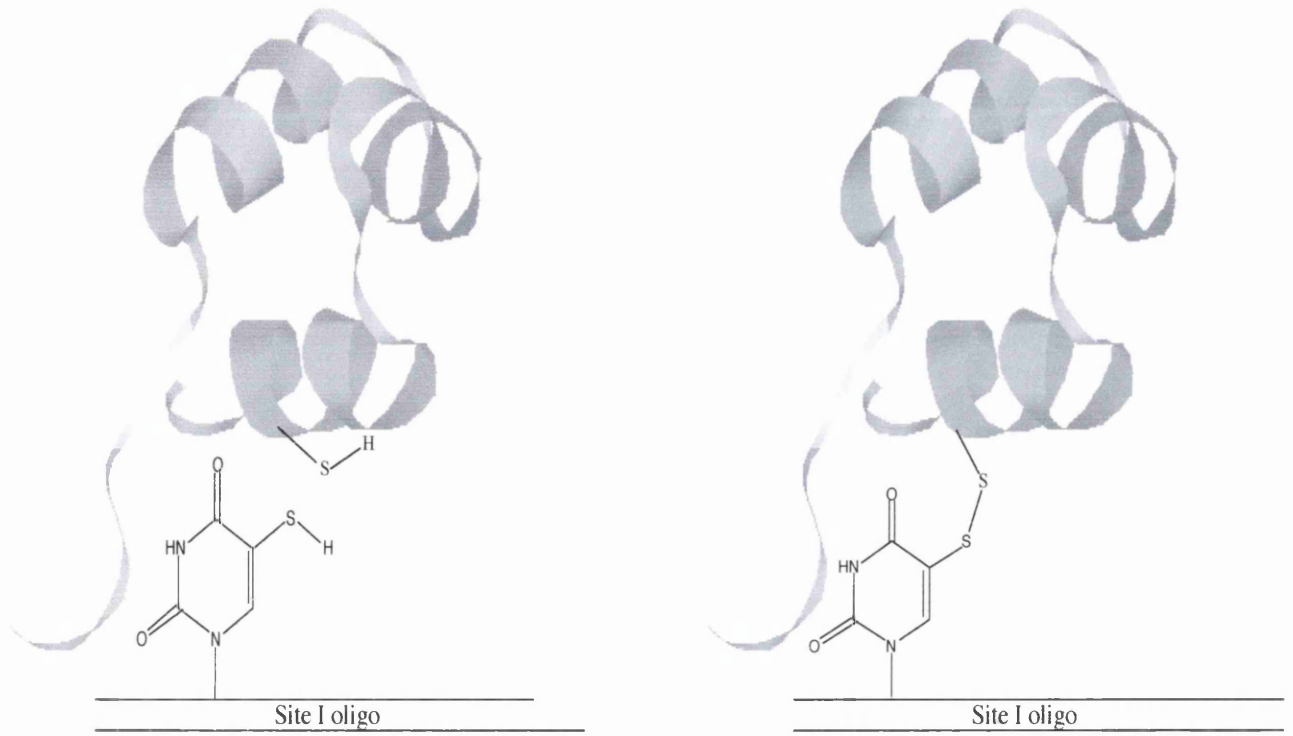
These results indicate that the mutants A171C, S173C and K177C are able to bind to binding site I with the bottom strands B1 and B3.

An apparent correlation between the amount of DTT and the extent of binding was found. At higher concentrations of DTT the binding was generally at higher levels. For the purposes of crosslinking this could be problematic, as increasing the amount of reducing agent in the binding reactions will discourage disulfide formation.

## **5.6 Crosslinking to mutant proteins**

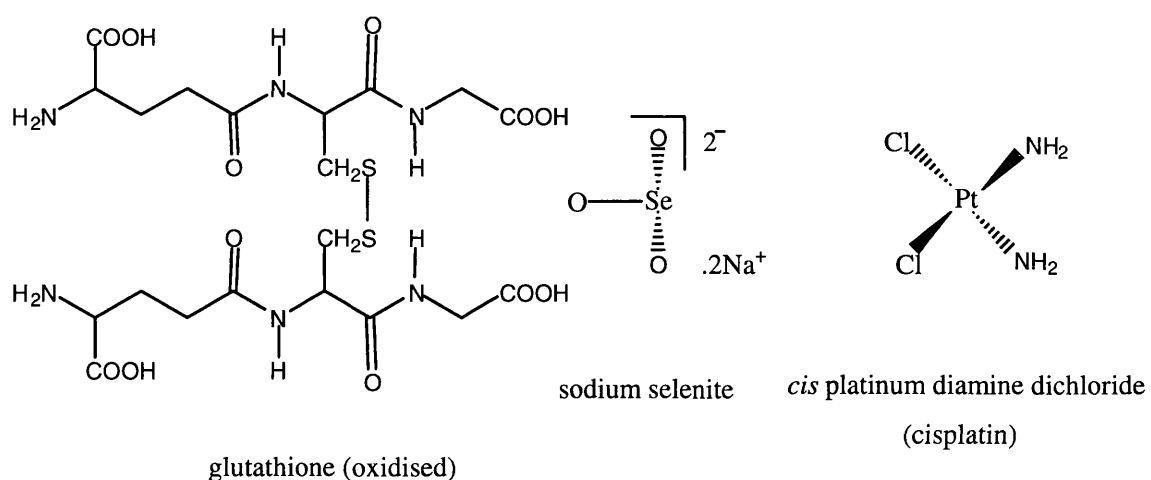
It was predicted that placing two thiol groups, one in the resolvase binding domain, and one in an oligonucleotide of the binding site, in close enough proximity should result in a disulfide bond being formed, thus forming a covalently linked protein-DNA complex (Figure 5.8).

For all crosslinking reactions (Section 3.42), a binding reaction was first carried out to enable the protein and DNA to come into close contact. The second part of the reaction involved the addition of oxidising agent to encourage the thiols to



**Figure 5.8 Crosslinking of the DNA binding domain**

Illustration of disulfide formation between a cysteine thiol in the DNA binding domain and a thiol-containing base in a site I oligonucleotide.



**Figure 5.9 Agents used in crosslinking**

Diagram illustrating the oxidising and crosslinking agents used to link C-5 thiols in DNA and cysteines in the recognition helix of Tn3 resolvase. Air oxidation was also used.



form disulfides. The oxidising and crosslinking reagents tested are illustrated in Figure 5.9. The products of the crosslinking reactions were run on 10% Tris-glycine gels containing 0.1% SDS. The gels were dried, and the products visualised using autoradiography.

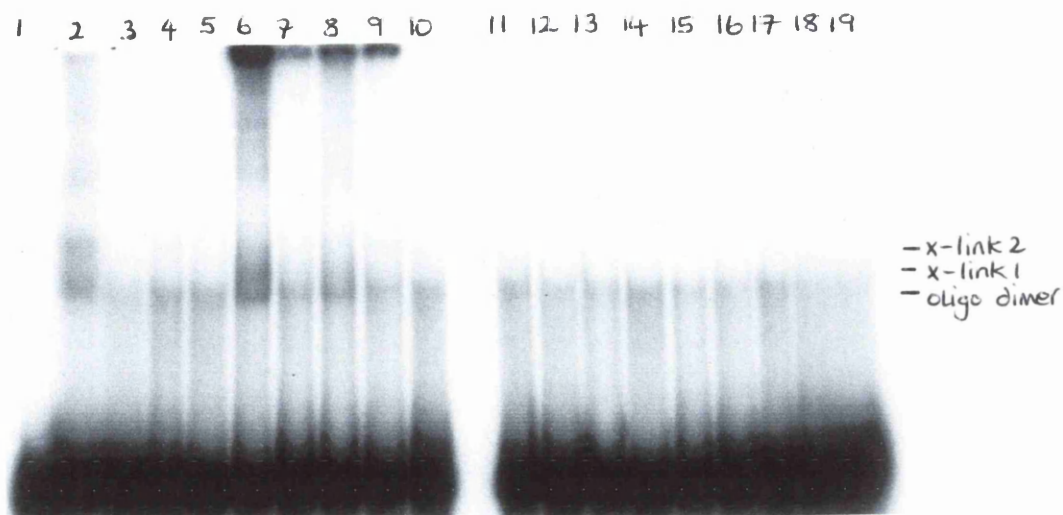
In most of the crosslinking assays (Figure 5.10-5.16) there is a weak band approximately half way between the top of the gels and the unbound oligonucleotides. This band is due to products of oligonucleotide oxidation only, as it is not sensitive to proteinase K (see right half of gels). This band is probably due to the thiol-containing oligonucleotides being oxidised. A small proportion of the oligonucleotides may form disulfides with each other, yielding dimeric species.

In lane 1 of all the crosslinking gels, which contains binding buffer and oligonucleotide only, there are varying amounts of this oxidised oligonucleotide species. The amount depended on how long the oligonucleotide solutions had been stored prior to use in the crosslinking assays. If the DTT in the buffer was depleted (initial thiol-containing oligonucleotides were in 1 mM DTT) then oxidation may occur over time. The oligonucleotides that had no dimer present were used immediately in the crosslinking reactions.

In some of the gels there are proteinase-K sensitive bands in the wild-type resolvase controls. In all cases, the presence of the cysteine in the mutants increased the amount of crosslinking, indicating that the thiol modification is causing more crosslinking than with wild-type protein.

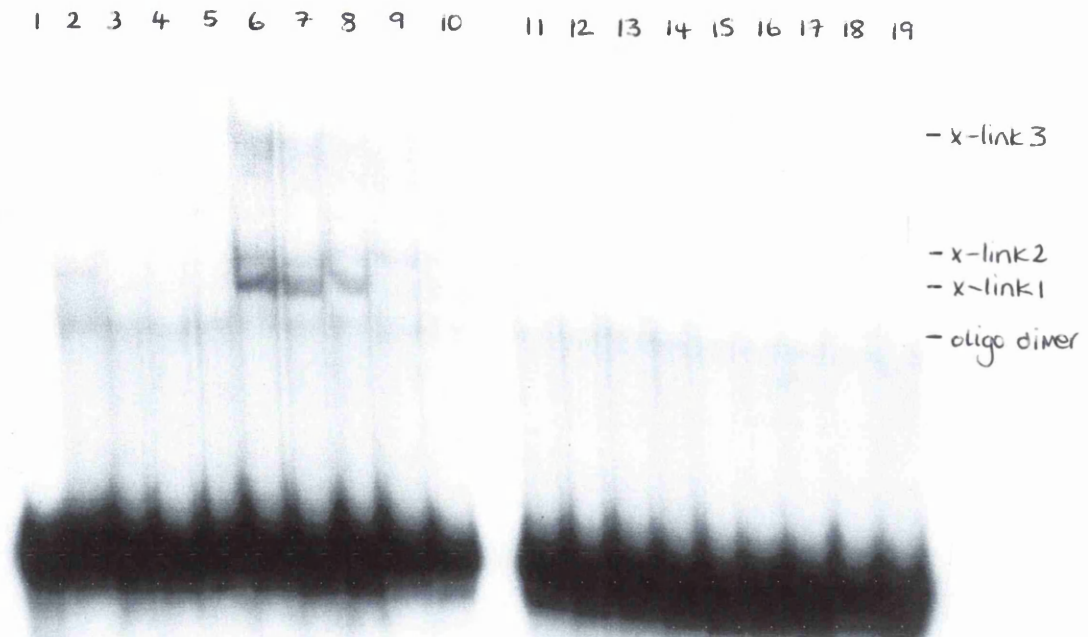
### **5.7 Crosslinking of double stranded oligonucleotide with modification at position B1**

The oligonucleotide B1 has the first thymidine in the consensus TGT modified with a thiol group, in the 48 nt bottom strand.



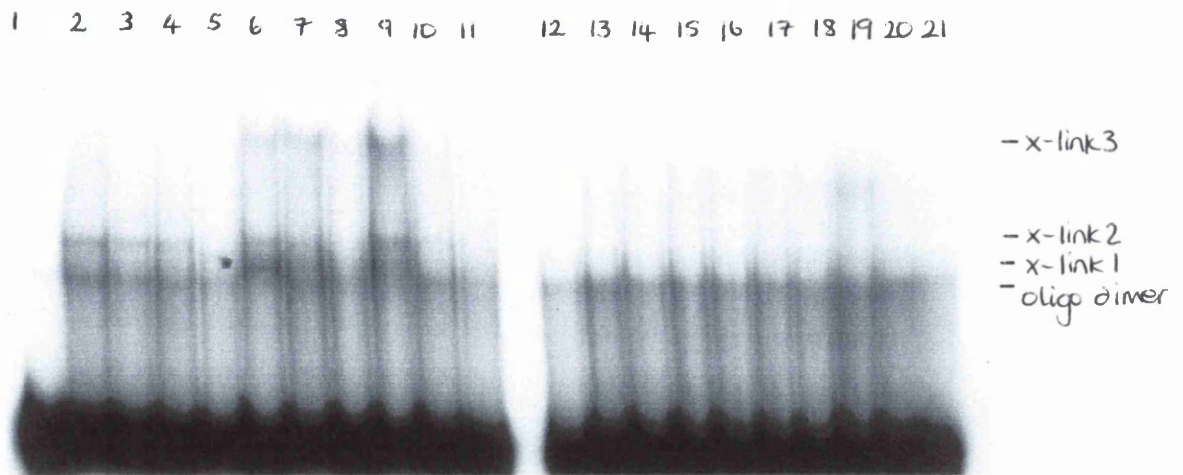
**Figure 5.10 Crosslinking of double stranded oligonucleotide with modification at position B1 to A171C, R172C and S173C with glutathione oxidation**

- Lanes 1 - No resolvase (binding buffer)  
 2 - 16-fold dilution of wild-type resolvase  
 3 - Undiluted A171C resolvase  
 4 - 2-fold dilution of A171C resolvase  
 5 - 4-fold dilution of A171C resolvase  
 6 - Undiluted R172C resolvase  
 7 - 2-fold dilution of R172C resolvase  
 8 - 4-fold dilution of R172C resolvase  
 9 - Undiluted S173C resolvase  
 10 - 2-fold dilution of S173C resolvase  
 11-19 - as lane 2-10 + proteinase K



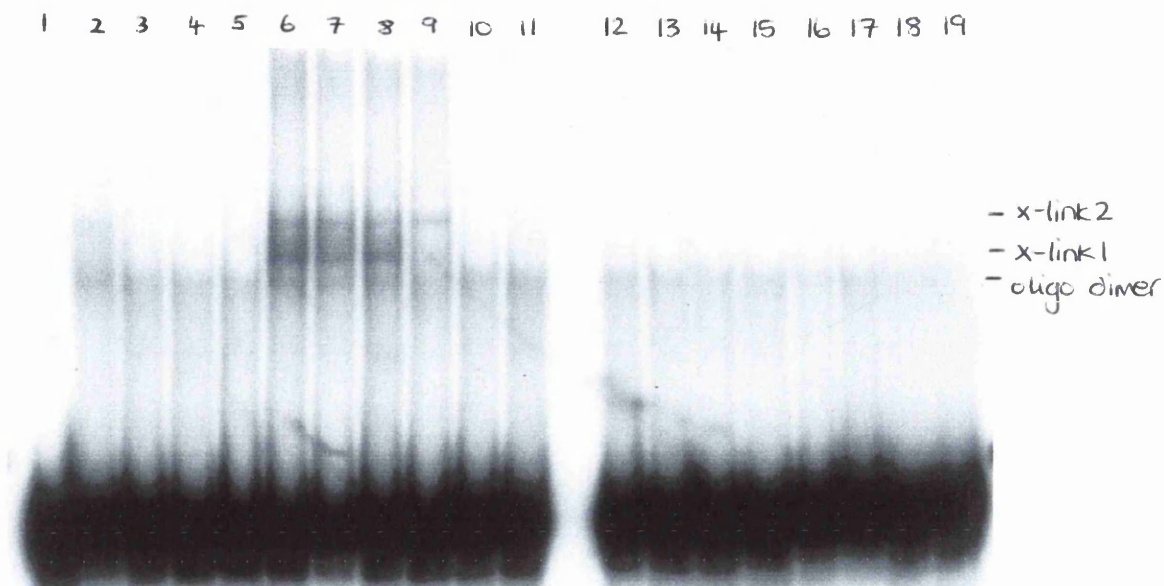
**Figure 5.11 Crosslinking of double stranded oligonucleotide with modification at position B1 to A171C, R172C and S173C with selenite oxidation**

- Lanes 1 - No resolvase (binding buffer)  
 2 - 16-fold dilution of wild-type resolvase  
 3 - Undiluted A171C resolvase  
 4 - 2-fold dilution of A171C resolvase  
 5 - 4-fold dilution of A171C resolvase  
 6 - Undiluted R172C resolvase  
 7 - 2-fold dilution of R172C resolvase  
 8 - 4-fold dilution of R172C resolvase  
 9 - Undiluted S173C resolvase  
 10 - 2-fold dilution of S173C resolvase  
 11-19 - as lane 2-10 + proteinase K



**Figure 5.12 Crosslinking of double stranded oligonucleotide with modification at position B1 to A171C, R172C and S173C with air oxidation**

- Lanes 1 - No resolvase (binding buffer)  
 2 - 16-fold dilution of wild-type resolvase  
 3 - Undiluted A171C resolvase  
 4 - 2-fold dilution of A171C resolvase  
 5 - 4-fold dilution of A171C resolvase  
 6 - Undiluted R172C resolvase  
 7 - 2-fold dilution of R172C resolvase  
 8 - 4-fold dilution of R172C resolvase  
 9 - Undiluted S173C resolvase  
 10 - 2-fold dilution of S173C resolvase  
 11 - 4-fold dilution of S173C resolvase  
 12-21 - as lane 2-11 + proteinase K



**Figure 5.13 Crosslinking of double stranded oligonucleotide with modification at position B1 to A171C, R172C and S173C with cisplatin oxidation**

- Lanes 1 - No resolvase (binding buffer)  
 2 - 16-fold dilution of wild-type resolvase  
 3 - Undiluted A171C resolvase  
 4 - 2-fold dilution of A171C resolvase  
 5 - 4-fold dilution of A171C resolvase  
 6 - Undiluted R172C resolvase  
 7 - 2-fold dilution of R172C resolvase  
 8 - 4-fold dilution of R172C resolvase  
 9 - Undiluted S173C resolvase  
 10 - 2-fold dilution of S173C resolvase  
 11 - 4-fold dilution of S173C resolvase  
 12-19 as lane 2-10 + proteinase K

Figure 5.10 shows crosslinking between B1 and A171C, R172C and S173C using oxidised glutathione as the oxidant. There are two faint proteinase K-sensitive bands in lane 6 (x-link 1 and x-link 2), corresponding to crosslinked complexes between R172C and B1. Although R172C showed very little binding to *res* and site I DNA in the binding assays (Section 4.14) it seems to have crosslinked in this assay. There is a lot of material stuck in the wells in lanes with the R172C crosslinking reaction. This material could be a result of aggregation of the R172C crosslinked complexes. The distance between the  $\beta$ -carbon of R172 and the C-5 of thymidine B1 (Table 4.1) was 7.71 Å, longer than the predicted distance (<5.8 Å) for a disulfide crosslink to be formed.

Figure 5.11 shows crosslinking between B1 and A171C, R172C and S173C using sodium selenite as the oxidising agent. Similarly to crosslinking with glutathione, there are crosslinked species with the mutant R172C as observed in lanes 6-8 (x-link 1). For all concentrations of R172C used, proteinase K-sensitive bands were observed. The amount of these complexes is around 5% of the total oligonucleotide. There is also a very weak band in lane 9 (x-link 2), so S173C has been crosslinked to B1. The distance between the  $\beta$ -carbon of R172 and the C-5 of thymidine B1 is 11.32 Å, so this reduction in crosslinking could be because the thiols are too far away.

Figure 5.12 shows crosslinking between B1 and A171C, R172C and S173C using air oxidation. Proteinase K-sensitive bands were observed in all the proteins assayed, although more crosslinked complexes were found with R172 (x link 1; lane 6 and 7 and R173 (x-link 2; lane 9). These 'xlink-2' bands have slightly different mobilities, probably due to the crosslinking to two different cysteine mutants. There are also crosslinked species between A171C and B1 (lanes 2 and 3) that appear to have the same mobility as the complexes in the wild-type control. The  $\beta$ -carbon atom of A171 is 12.87 Å away from the C-5 of B1.

Figure 5.13 shows crosslinking between B1 and A171C, R172C and S173C, using cisplatin. The mutant R172 in lanes 6-8, shows two crosslinked species (x-link 1 and x-link 2). There is also crosslinking between S173C and B1 (lane 9).

### **5.8 Crosslinking of double stranded oligonucleotide with modification at position B3**

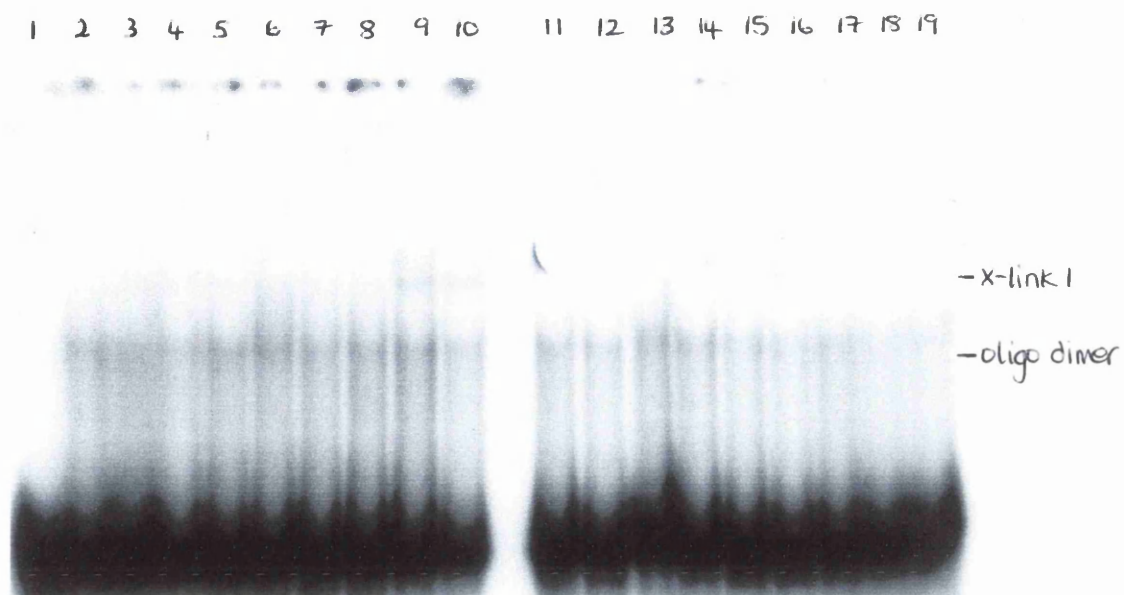
The oligonucleotide B3 has the third thymidine in the consensus TGT modified with a thiol group, in the 48 nt bottom strand. The crosslinking agents used on B1 oligonucleotide with the cysteine mutant proteins were the same as used for the B3 oligonucleotide. Covalent protein-DNA complexes were observed when air oxidation and cisplatin were used.

Figure 5.14 shows the crosslinking reaction of A171C, R172C and S173C with oligonucleotide B3 with air oxidation. The gel shows that crosslinking between S173C and B3 has occurred (see lanes 9 and 10). The band is faint, indicating that a very low amount of crosslinked product is present. Less than 5% of the total oligonucleotide added is covalently linked to the protein. The cysteine of S173C is the closest to the modified base of the three proteins used in this assay, being 6.31 Å away.

Figure 5.15 shows the crosslinking reaction of A171C, R172C and S173C with oligonucleotide B3 with cisplatin as the oxidant. Similar results were observed to Figure 5.14. A band corresponding to crosslinked complex was observed in lane 9 and 10 (xlink-1). Thus S173C has formed a covalent complex with oligonucleotide B3; the distance between B3 and S173C is 6.31 Å. More crosslinked complex was observed with cisplatin than with air oxidation.

### **5.9 Crosslinking of double stranded oligonucleotide with modification at position B5**

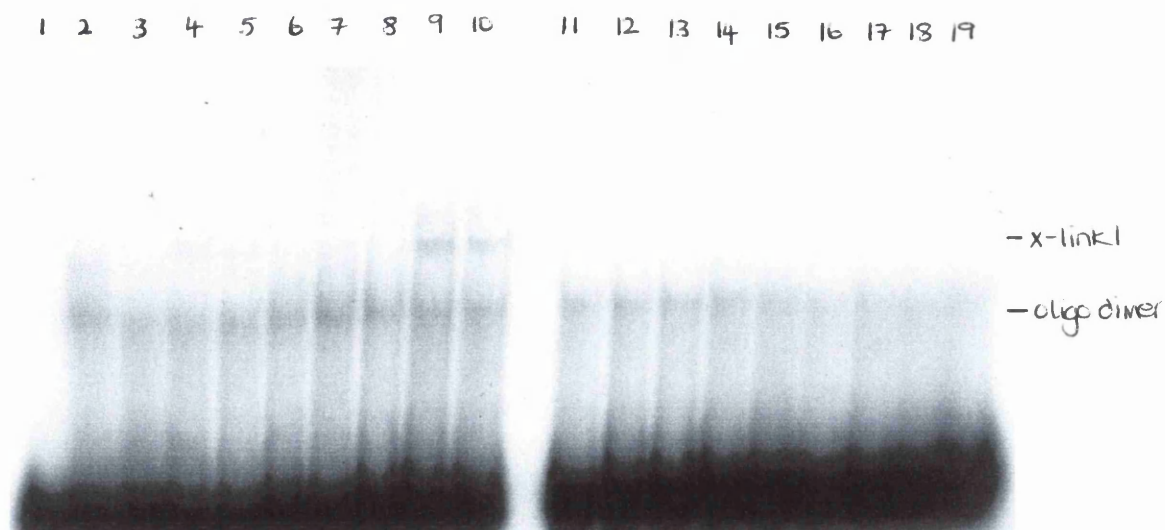
The double stranded site I oligonucleotide with the B5 thymidine modified in the bottom strand was used in crosslinking reactions with all of the cysteine mutants.



**Figure 5.14 Crosslinking of double stranded oligonucleotide with modification at position B3 to A171C, R172C and S173C with air oxidation**

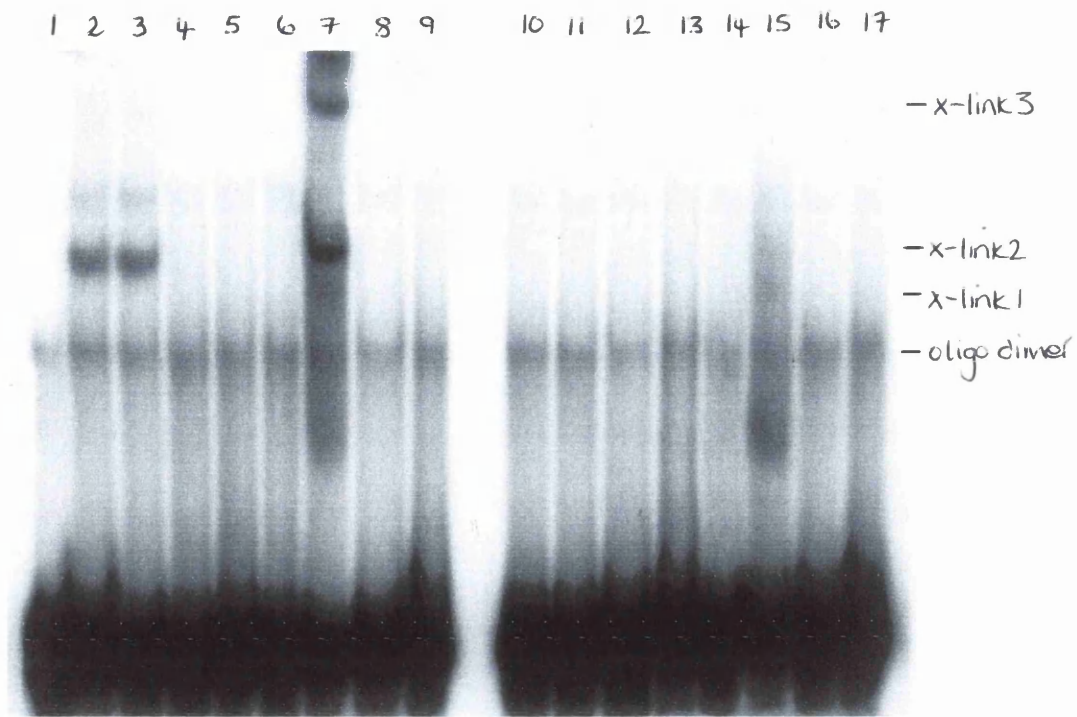
- Lanes 1 - No resolvase (binding buffer)  
 2 - 16-fold dilution of wild-type resolvase  
 3 - Undiluted A171C resolvase  
 4 - 2-fold dilution of A171C resolvase  
 5 - 4-fold dilution of A171C resolvase  
 6 - Undiluted R172C resolvase  
 7 - 2-fold dilution of R172C resolvase  
 8 - 4-fold dilution of R172C resolvase  
 9 - Undiluted S173C resolvase  
 10 - 2-fold dilution of S173C resolvase  
 11-19 - as lane 2-10 + proteinase K





**Figure 5.15 Crosslinking of double stranded oligonucleotide with modification at position B3 to A171C, R172C and S173C with cisplatin oxidation**

- Lanes 1 - No resolvase (binding buffer)  
 2 - 16-fold dilution of wild-type resolvase  
 3 - Undiluted A171C resolvase  
 4 - 2-fold dilution of A171C resolvase  
 5 - 4-fold dilution of A171C resolvase  
 6 - Undiluted R172C resolvase  
 7 - 2-fold dilution of R172C resolvase  
 8 - 4-fold dilution of R172C resolvase  
 9 - Undiluted S173C resolvase  
 10 - 2-fold dilution of S173C resolvase  
 11-19 - as lane 2-10 + proteinase K



**Figure 5.16 Crosslinking of double stranded oligonucleotide with modification at position B5 to A171C and Y176C with cisplatin oxidation**

- Lanes 1 - No resolvase (binding buffer)  
 2 - 32-fold dilution of wild-type resolvase  
 3 - 32-fold dilution of S10N resolvase  
 4 - Undiluted A171C resolvase  
 5 - 2-fold dilution of A171C resolvase  
 6 - 4-fold dilution of A171C resolvase  
 7 - Undiluted Y176C resolvase  
 8 - 2-fold dilution of Y176C resolvase  
 9 - 4-fold dilution of Y176C resolvase  
 10-17 - as lane 2-9 + proteinase K

The best result was obtained with the Y176C mutant as shown in Figure 5.16 (x-link 1, x-link 2 and x-link 3, as seen in lane 7). The band x-link 1 represents Y176C crosslinked to B5. The distance between the  $\beta$ -carbon of Y176 and the C-5 of thymidine B5 is 9.94 Å. The more retarded band crosslink 2 is a perhaps due to an extra monomer being crosslinked as in Figure 5.18. The very retarded species crosslink 3 is probably due to crosslinked, aggregated complexes as it is sensitive to proteinase K, and was rather stable to 0.1% SDS, depending on the amount of aggregation and the amount of SDS actually present in the gels.

In some of the crosslinking gels, there was a retarded band in the wild-type control reactions that was sensitive to proteinase K treatment. As wild-type control contains no cysteine residues, this band is not due to disulfide formation between protein and DNA, but due to crosslinking by some other method. The experiments were repeated using the mutant S10N resolvase that lacks the serine 10 nucleophile (M. Stark, pers. comm.). This should test whether the band in the wild-type control represents the oligonucleotide covalently joined to the protein *via* the nucleophilic serine of resolvase's active site. The fact that this band is also present in S10N crosslinking reactions (Figure 5.16) indicated that crosslinking to the wild-type protein was not specifically at serine 10. For the cisplatin gels the reaction (Figure 5.17) could account for these bands, but for the other reagents it is not obvious what the covalent species are.

The equivalent gels for B1, B3 and B5 using the same proteins that gave crosslinked complexes, with no oxidising agents present were run. There were no bands present in any of the lanes, except oxidised oligonucleotide. The absence of bands in the wild-type control, indicates that the wild-type bands require the presence of oxidants and the bands seen with the mutant proteins are real.

The smeary retarded band (xlink 3) in Figures 5.11, 5.12, 5.13 and 5.16 could be due to large protein aggregates that are insensitive to SDS present in the gel. When the reagent cisplatin was used there were marked bands in the

wild-type controls. This could be due the cisplatin reacting with guanine in the DNA and with a nearby reactive amino group, e.g. lysine at position 177 (Figure 5.17). X-link (2) could be due to extra protomers being oxidised as in Figure 5.18.

## 5.10 Discussion

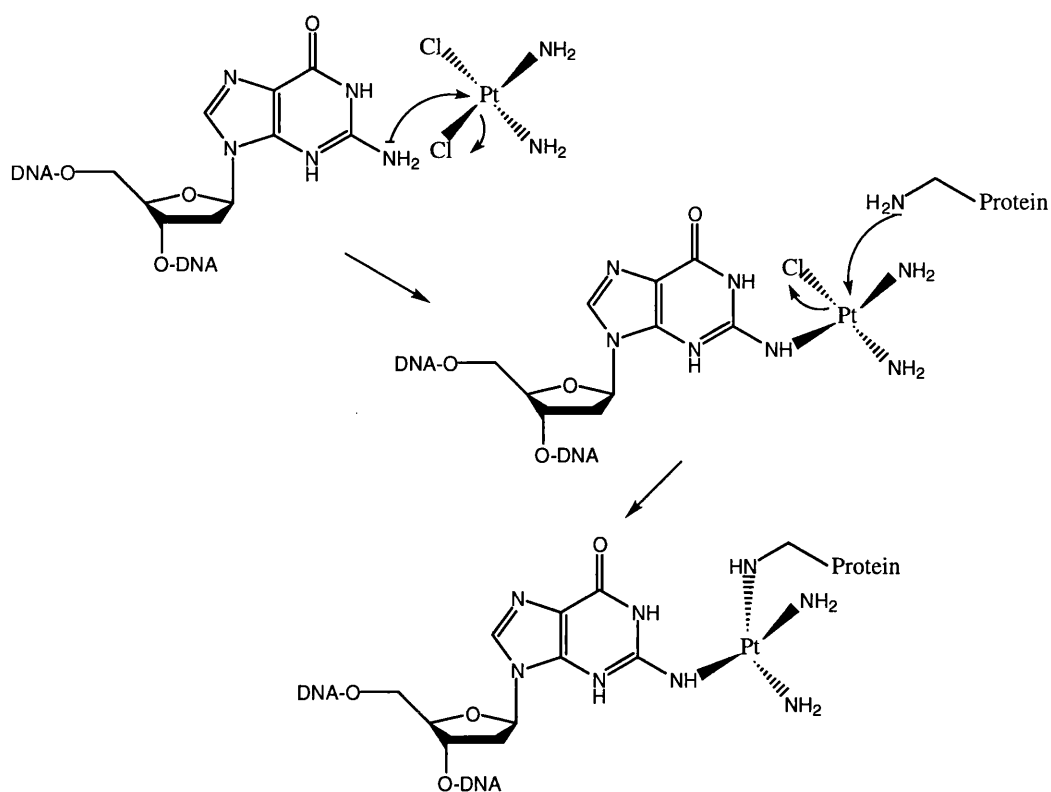
Initial expectations for levels of crosslinking were unknown. Unfortunately, the levels of crosslinking obtained were lower than hoped for.

The two major components required for the disulfide crosslinking approach, the novel cysteine mutants (Chapter 3) and the thiol-containing oligonucleotides (Chapter 4), were produced during this work.

The evidence for crosslinking is the bands observed on the denaturing SDS Tris-glycine gels (Figure 5.10-5.16) that were covalent protein-DNA complexes, sensitive to proteinase K treatment, suggesting that the bands did not arise from DNA/DNA crosslinking; the disappearance of the crosslinked complexes on treatment with proteinase K; and the absence of any protein-DNA adducts when the crosslinking reactions were carried out in the absence of oxidants, indicating that the crosslinking was dependent on the thiols and being oxidised to sulfides.

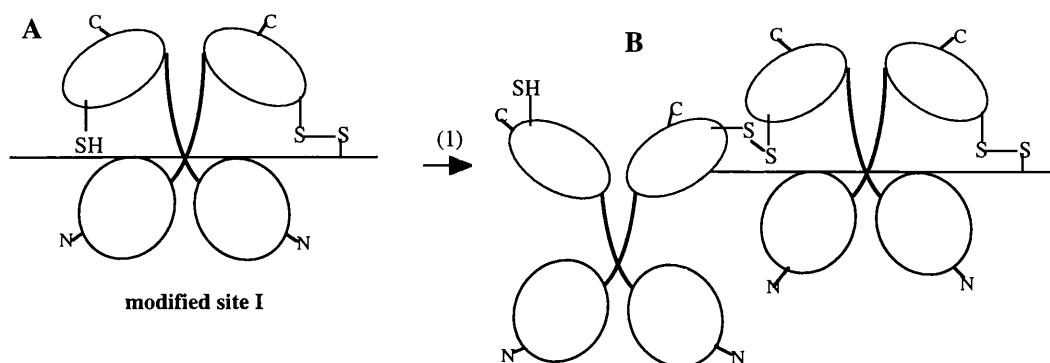
The formation of thiol-dependent crosslinks was further confirmed by the following; the amount of crosslinking was increased when cysteine mutants and thiol-containing oligonucleotides were used, as compared to crosslinking with the wild-type protein and unmodified oligonucleotides.

The mutation of important residues in the recognition helix was expected to diminish binding to the DNA. However the mutants A171C, S173C and K177C were still able to bind to both wild-type and modified site I oligonucleotides, although more weakly than wild-type resolvase. The pKa of a cysteine thiol is about 9.5, meaning that at pH 9.5 50% of the thiol would be ionised. As the



**Figure 5.17 Additional bands in cisplatin reactions**

The cisplatin could crosslink a guanine base to a nearby lysine, as shown above.



1. Binding and crosslinking reaction

**Figure 5.18 The additional band in crosslinking gels**

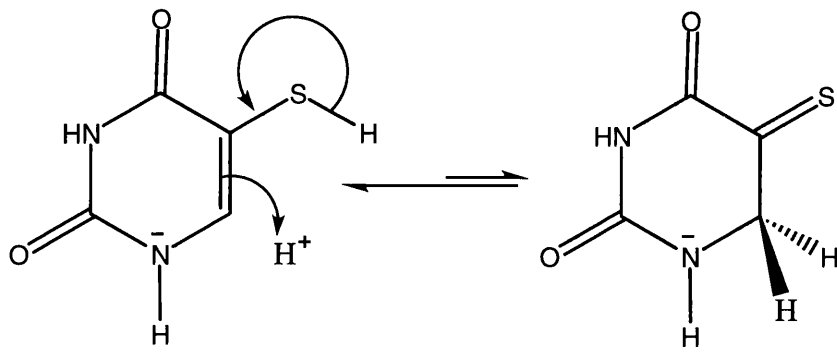
Species **A** indicates a disulfide between a modified strand and one dimer of cysteine mutant. In **B** an additional oxidised species may be formed. N denotes the N-terminal domain of resolvase and C indicates the C-terminal binding domain.

the cysteine thiols may have been negatively charged, which may cause a repulsion between it and the similarly ionized thiol present in the oligonucleotide. This factor would contribute to the reduced binding that was observed. In order to generate more of the protonated thiol species, the crosslinking reactions were carried out in Tris buffer at pH 7.5. However, this did not cause a greater amount of crosslinking.

Another factor that could contribute to the reduced levels of crosslinking is the amount of thiol present in the modified oligonucleotides themselves. There could be decreased amount of thiol present due to incomplete deprotection with  $\beta$ -mercaptoethanol (Section 5.4). From the alkylation experiment (Figure 5.3), it was calculated that around 90% of the free thiol was alkylated, meaning that 10% of the oligonucleotides still contained the DNP group. The thiol within the oligonucleotides may have been lost or chemically altered and this would reduced the amount available for crosslinking.

The type of thiol modification used, replaces the methyl group of thymidine with a thiol moiety. Within the pyrimidine ring system of thymidine there is an  $\alpha$ - $\beta$ -unsaturated ketone which activates the 6-position and deactivates the 5-position. Thus delocalisation of the thiol lone pair throughout the pyrimidine ring renders the thiol less nucleophilic, and forms a thione group as shown in Figure 5.19. It was hoped that the equilibrium between these two states would be such that any tautomerisation from the thione to the thiol would result in disulfide bond formation to the thiol from the protein. The thiol form should be stabilised due to conjugation in the pyrimidine ring. The thione moiety has a characteristic UV absorbance that was not detected (Chapter 2).

The cysteine thiol present in the mutant resolvases and the thiol in the oligonucleotides may simply have been too far apart for high levels of crosslinking to be achieved. The distances taken from the co-crystal structure (Table 4.1), suggested that the two reacting thiols in Y176C and B3 were close



**Figure 5.19 Thiol-thione tautomerisation**

The thiol at the 5-position of the pyrimidine ring can exist in two forms, the free thiol and the thione conformation.



enough ( $<5.6 \text{ \AA}$ ) to form disulfides. The distances were longer for all the other possible combinations of modified oligonucleotide and cysteine thiol. Perhaps the thiols did not form close enough interactions to be crosslinked.

In order to determine whether the retarded band present in the binding gels (Figure 5.6 and 5.7) contained a covalent disulfide linkage, a 2D gel was run, with the normal binding gel in the first dimension and an SDS gel in the second dimension. This should cause dissociation of any non-covalent complexes, while maintaining the integrity of the covalent complex. Unfortunately the result of the gel was inconclusive (data not shown).

Although the site I DNA used in the crosslinking had nearly identical binding motifs at each end, there was a noticeable difference in the crosslinking efficiency of each half site. All the top strand modifications were in the left hand side of site I, whereas the bottom strands were at the right half. All attempts to form crosslinks with the top strand-modified oligonucleotides **T1**, **T3** and **T4** failed, protein-DNA complexes being isolated for bottom strand **B1**, **B3** and **B5** modified oligonucleotides only. It has previously been shown that the left half of site II binds more strongly to resolvase (Blake *et al.*, 1995), and previous work indicated that the order of half-site binding affinities, from strongest to weakest, were  $\text{III-L} \approx \text{II-L} > \text{I-R} > \text{III-R} \approx \text{I-L} > \text{II-R}$  (Abdel-Meguid *et al.*, 1984).

When performing the spatial analysis (Table 4.1) of the distances between the  $\beta$ -carbon atom of the amino acid to be mutated and the 5-position of the thymidines in the region, it was obvious that there were differences in the distances between the protein and the bases at each end of the crystallographic site I DNA (Yang and Steitz, 1995). The distances between the protein and the DNA at one side of the oligonucleotides were marginally smaller than the equivalent contacts at the other side. However, as the oligonucleotide in the co-crystal structure (Yang and Steitz, 1995) is symmetrical in sequence, it is impossible to say whether the side that is closest is relevant to the crosslinking results obtained here.

The lack of crosslinked complexes with modifications at the left hand side of site I DNA may indicate that a stronger interaction occurs at the right side of the protein-DNA complex, so introducing modifications there has a greater effect. The right end monomer may bind first to the DNA in a strong interaction. Once this right side monomer-protein DNA complex is formed a second monomer may bind at the left side of the DNA. Once dimerisation is complete, the right end which bound first may be held more tightly to the DNA than the left side.

Position 176 is the closest of all the mutated residues to the DNA. It forms both hydrophobic and electrostatic interactions with the bases, and in the wild-type protein the tyrosine residue is essential for maintenance of the recognition helix. Previously it was shown that the mutant Y176C is defective in binding and has reduced recombination activity. The presence of crosslinked complexes shows that covalent bonds have been formed between the protein and the site I DNA.

Similarly R172 has many important interactions with the DNA, bound to *res* DNA very weakly, and had no full recombination activity. However it did form crosslinks with double stranded oligonucleotides containing thiol modifications at positions B1 and B3.

## **Chapter 6: Investigation of the stereochemical properties of Tn3 resolvase**

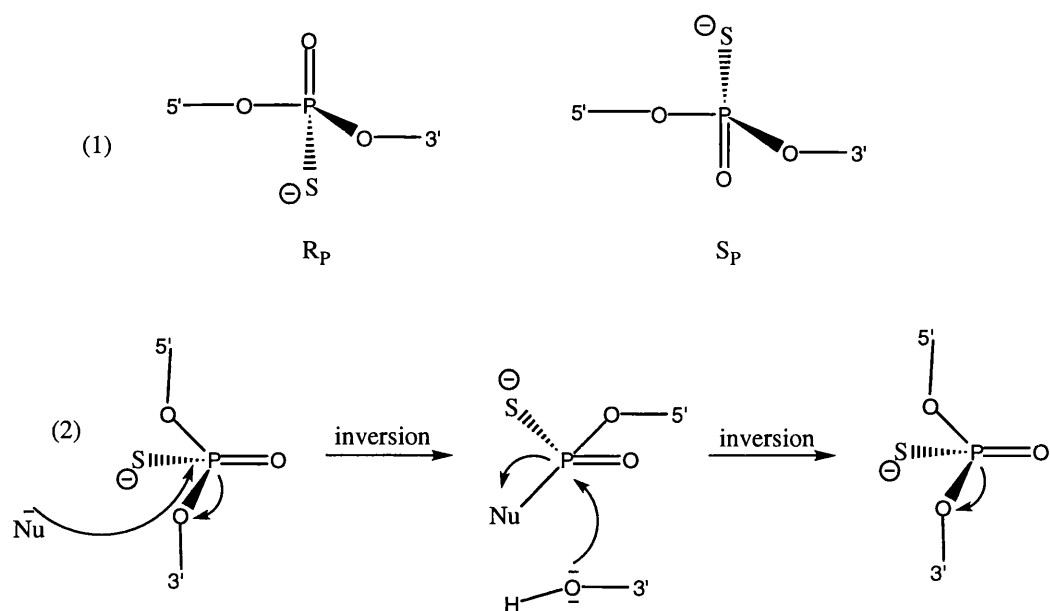
## 6.1 Introduction

The stereochemical course of the strand exchange reaction involved in site specific recombination by the Tn3 resolvase will be determined by making reaction substrates that contain phosphorothioates. The number of catalytic steps can be inferred from the stereochemistry of a phosphorothioate. For one-step (or any odd number of) transesterifications there will be inversion of the chirality, and for a two step (or any even number) mechanism the chirality of the phosphate will be retained.

The replacement of one non-bridging oxygen atom of a DNA phosphodiester by a sulfur atom in the phosphorothioate is a conservative change, as the negative charge of the phosphate group is retained, and the size of the sulfur atom is only slightly larger than the oxygen atom. The change does, however, make the phosphorus centre chiral; that is, there are two diastereoisomeric forms of a phosphorothioate linkage (Figure 6.1). Phosphorothioates have been shown to possess greater resistance to enzymatic cleavage at the internucleoside linkage, and the sulfur atom is more nucleophilic than oxygen.

Use of phosphorothioates in studying protein/DNA interactions has been extensive (reviewed in Verma and Eckstein, 1998). Site-specific incorporation of phosphorothioates has been used to study reaction mechanisms of restriction enzymes such as *EcoRI* (Connolly *et al.*, 1984) and *EcoRV* (Grasby and Connolly, 1992), HIV integration (Engelman *et al.*, 1991) and site-specific recombination (Kitts and Nash, 1988a). In addition, phosphorothioates have been used to identify oxygen atoms responsible for or involved in the interaction of specific phosphate groups with proteins *via* incorporation of R<sub>p</sub> and S<sub>p</sub> linkages at the recognition sites for *EcoRI* (Kurpiewski *et al.*, 1996) and *EcoRV* (Thorogood *et al.*, 1996).

The phage Mu system was studied using a chiral phosphorothioate to work out the number of steps at the strand transfer step, a transesterification reaction



**Figure 6.1 The enantiomeric configurations of a phosphorothioate**

(1) The sulfur atom renders the phosphodiester of the internucleotide linkage chiral. The two enantiomeric conformations are indicated above as  $S_p$  and  $R_p$ . (2) shows the nucleophilic attack at a phosphorus centre each attack being equivalent to an inversion of the stereochemistry of that phosphorus.

(Mizuuchi and Adzuma, 1991). The preference for R or S phosphorothioates at the transesterification reactions, catalysed by Tn10 transposase (Kennedy *et al.*, 2000) was also investigated. Phosphorothioates were used to study the mechanism of site-specific recombination in phage  $\lambda$  (Kitts and Nash, 1988b). Substitution of phosphorothioate for phosphate in one strand of the recombination site resulted in a block in recombination, leading to accumulation of Holliday structures.

*EcoRV* restriction endonuclease, acting on a chiral phosphorothioate, shows specificity for one (the R<sub>p</sub>) diastereoisomer; that is, *EcoRV* only cleaves oligonucleotides that contain a R<sub>p</sub> phosphorothioate (Grasby and Connolly, 1992).

This selectivity of *EcoRV* is used here to study the stereochemistry of the resolvase strand exchange reaction, which requires the breakage and religation of phosphodiester bonds in the DNA backbone. By introducing a phosphorothioate and an *EcoRV* site at the site in the DNA where resolvase cleaves and religates, it should be possible to determine the stereochemistry of strand exchange, and hence the number of catalytic steps during strand exchange. The currently proposed mechanism involves two catalytic steps; that is two nucleophilic steps, the first involves nucleophilic attack of the crossover site by ser10 of resolvase, forming the covalent protein-DNA complex. The second nucleophilic attack occurs *via* the 3' hydroxyl from the opposite strand, leading to recombinant products (Figure 1.16).

## 6.2 Construction of modified substrates

For  $\lambda$  integrase, the phosphorus of the phosphodiester that is cleaved is attached to the 3' end in the intermediate, whereas with resolvase the phosphorus is on the 3' end. The nature of the  $\lambda$  int intermediate allowed the use of a strategy (Kitts and Nash, 1988a) to characterize the phosphorothioate stereochemistry, by digestion of the DNA with nucleases which are specific to either the R or S stereoisomers of the phosphorothioate. If the phosphorus 5' of the

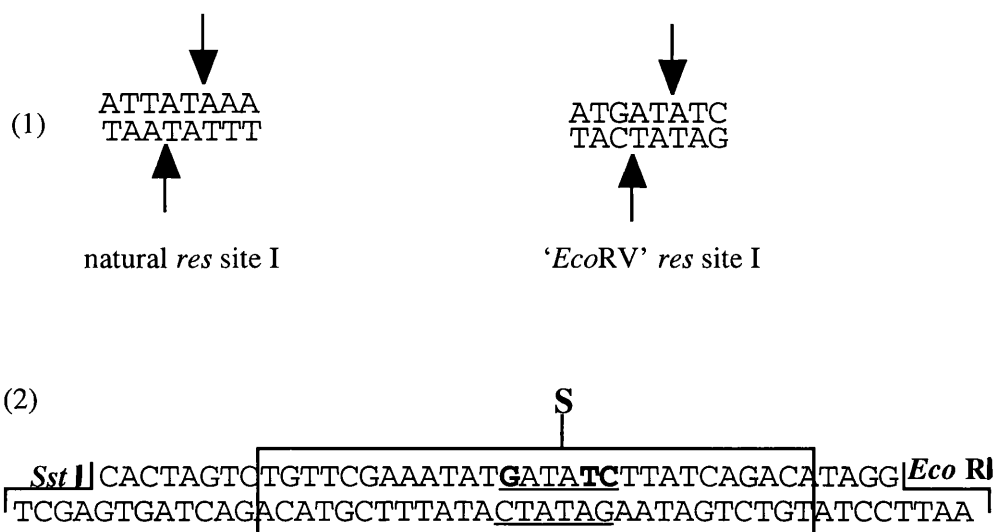
phosphorothioate is labelled, the phosphorothioate refractory to cleavage can be detected as a labelled dinucleotide. The nature of the resolvase intermediate precludes the use of this strategy.

With  $\lambda$  integrase, the 5'  $^{32}\text{P}$  label and the phosphorothioate remain on the same nucleotide before and after strand exchange. With Tn3 resolvase, strand exchange joins the left half of one site I in *res* to the right hand site of the equivalent *res* after strand exchange if the phosphorothioate and the label were at the same site I, the recombinant circle would either only be labelled or contain a phosphorothioate (not both).

An *EcoRV* site can be created at the centre of site I, by mutating 3 bp of natural site I, such that resolvase and *EcoRV* both cut the same phosphodiester. This is a feasible strategy because both *EcoRV* and resolvase cleave a phosphodiester after an AT dinucleotide at their recognition sequences.

A phosphorothioate would be introduced at the scissile bond using oligonucleotides. The substrate would be resolved using Tn3 resolvase. Then *EcoRV* would be used to determine the phosphorothioate chirality in the product (see Section 6.1). This approach requires that the recombinant site also contains an *EcoRV* site, so the partner site had to be mutated as well. The natural *res* binding site I was mutated at 3 bp to create an *EcoRV* site (Figure 6.2.1). A modified oligonucleotide (pMSRV) of the same sequence, with a chiral phosphorothioate at the centre of site I (Figure 6.2.2) was also made (Section 3.25).

The resolvase should change the configuration of the phosphorothioate or not, depending on how many steps the reaction proceed in, i.e. ( an odd or even number). The configuration at the centre of site I in the cleaved intermediate is illustrated in Figure 6.3.

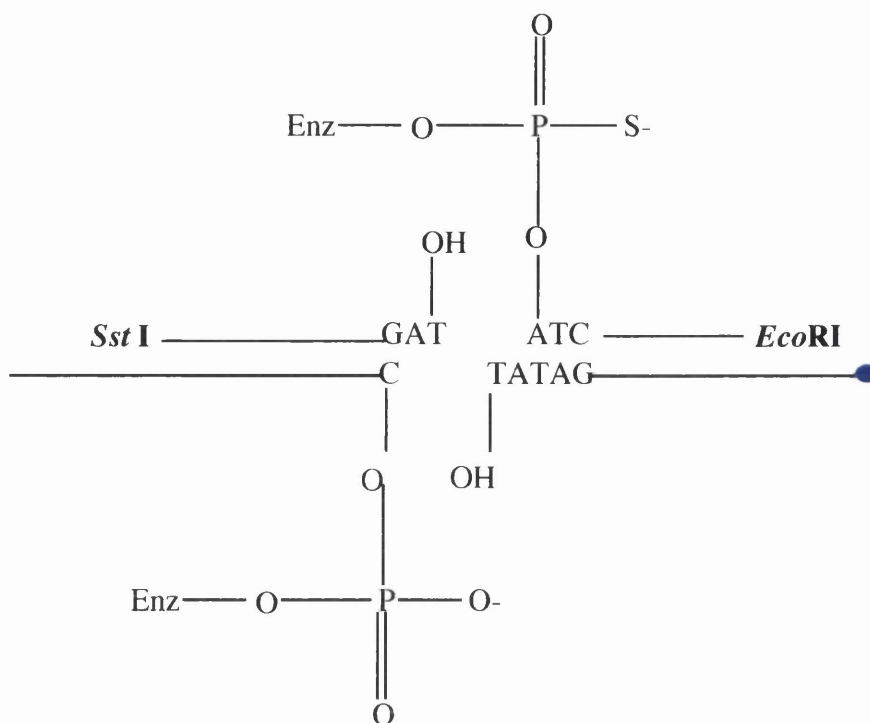


**Figure 6.2 Sequences of the middle of site I of *res***

(1) The natural palindromic sequence is shown on the left. The '*EcoRV*' site I on the right has 3 bp changed from the natural *res* site I, creating the recognition sequence for *EcoRV*. The site was derived from pMS53 (M. Burke; Figure 6.5). The arrows indicate the positions where resolvase cleaves the DNA. In the '*EcoRV*' site I, the cleavage by resolvase in the top strand is at the same position as cleavage by *EcoRV*.

(2) shows the clonable *SstI-EcoRI* site I oligonucleotide containing an *EcoRV* site at the centre of the site. The recognition sequence for *EcoRV* is underlined. The top strand 40-mer contains a phosphorothioate (S) between the TA as indicated. The bases mutated from wild-type are indicated in bold.





**Figure 6.3 The configuration at the crossover site of the cleaved intermediate**

Illustration of the configuration of the phosphorothioate at the central AT of site I, once the strands have been cleaved. Enz = resolvase. The radiolabelled phosphate is shown as a blue dot.

The construction of pMSH2 is shown in Figure 6.4. The *res* site in pMSH2 that has an *EcoRV* site at the centre of the crossover site will be modified with the phosphorothioate-containing oligonucleotide, and the DNA will be reconstituted into a supercoiled substrate (Section 3.36), radiolabelled with  $^{32}\text{P}$  at the *EcoRI* site (Section 3.16). This will ensure that the  $^{32}\text{P}$  goes to the same place as the top strand phosphorothioate on resolution (Figure 6.5).

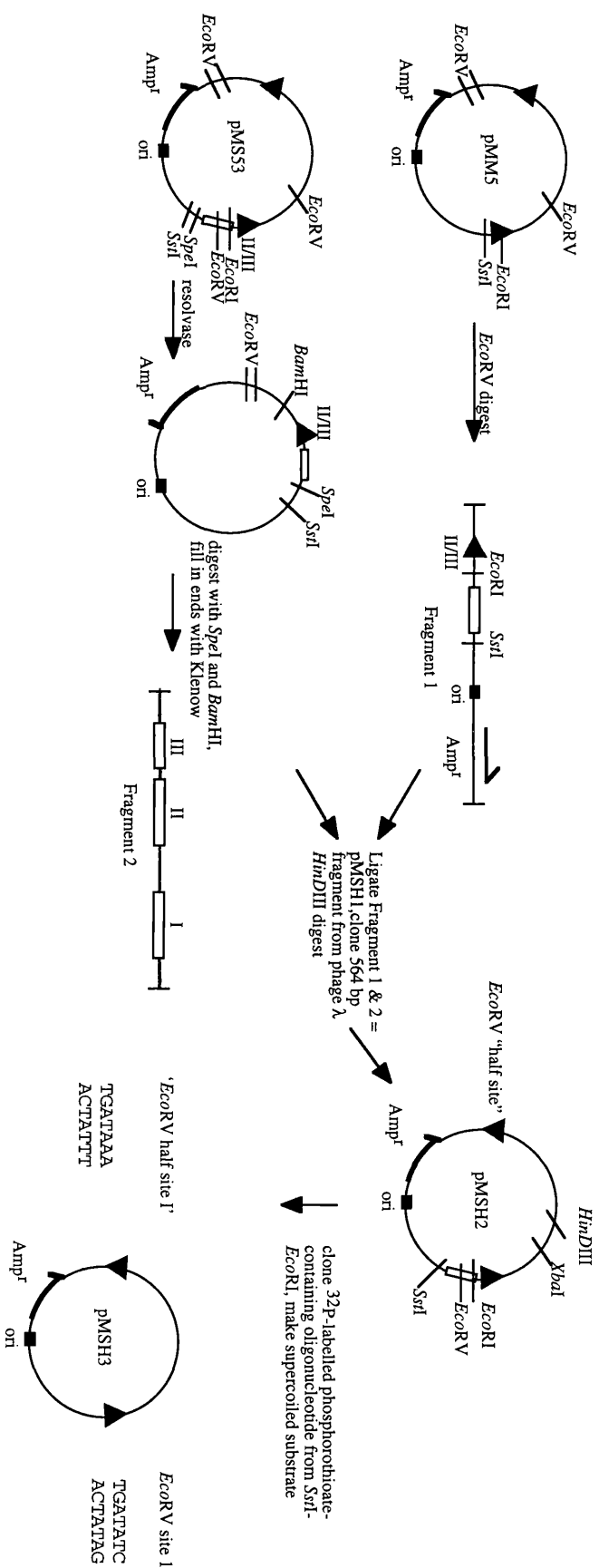
The plasmid pMS53 contains two *res* sites in direct repeat. One is a wild-type site and the other has an *EcoRV* site at the centre of site I. The modified *res* sites in pMSH2 contain an *EcoRV* site at the centre of one site I and an "*EcoRV* half site" at the other, so that the resolution product also contains an *EcoRV* site.

### 6.3 Resolution of pMSH2

Once the substrate pMSH2 was constructed, it was important to check that the resolution product contained an *EcoRV* site in the small circle. The resolution reaction of pMSH2 is shown in Figure 6.6. Both wild-type protein and D102Y (hyper-reactive mutant; Arnold, 1998) were used to resolve the substrates containing the mutated *res* sites. There is greater resolution with the D102Y mutant than with the wild-type protein (Figure 6.6 compare lanes 9 and 2). The digestion with *Bgl*III yields a linear 2.4 kbp fragment and a 800 bp supercoiled fragment. When the resolution products were cut with *EcoRV*, a linear 800 bp product and a supercoiled 2.4 kbp product were observed. Therefore the *EcoRV* site is in the smaller circle on resolution, as predicted.

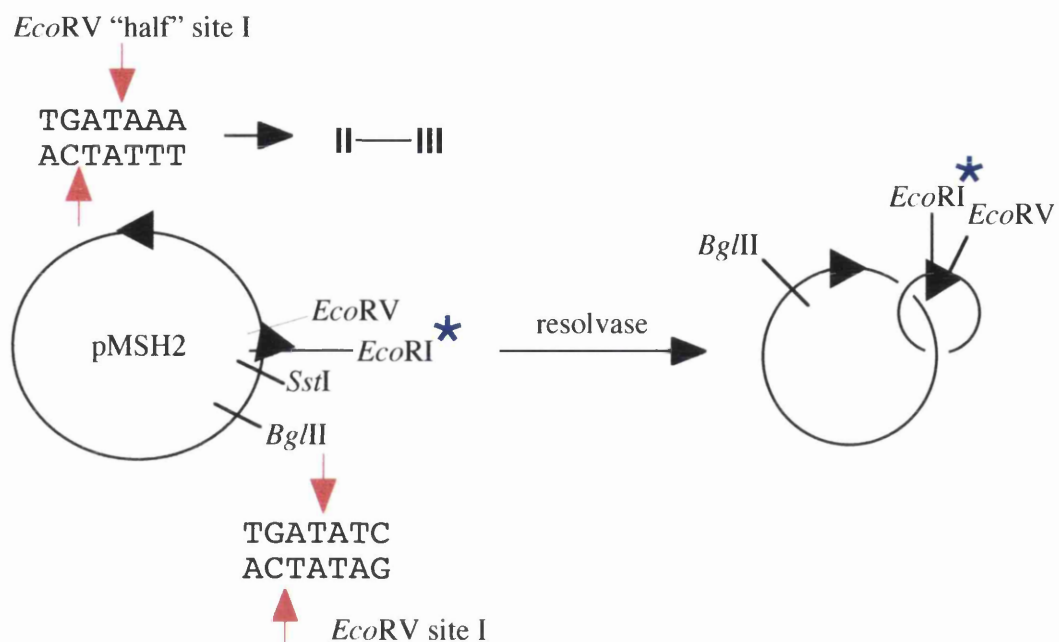
### 6.4 *EcoRV* titration of pMSH3

The phosphorothioate-containing  $^{32}\text{P}$  labelled oligonucleotides were ligated into pMSH2 (McIllwraith, 1995), yielding pMSH3 as the recombination substrate. The results of an *EcoRV* titration of pMSH3 are shown in Figure 6.7. The experiment was carried out in duplicate (set A and B) and the resulting intensity of the bands were determined by quantitation of the phosphoimage using the MacBas program.



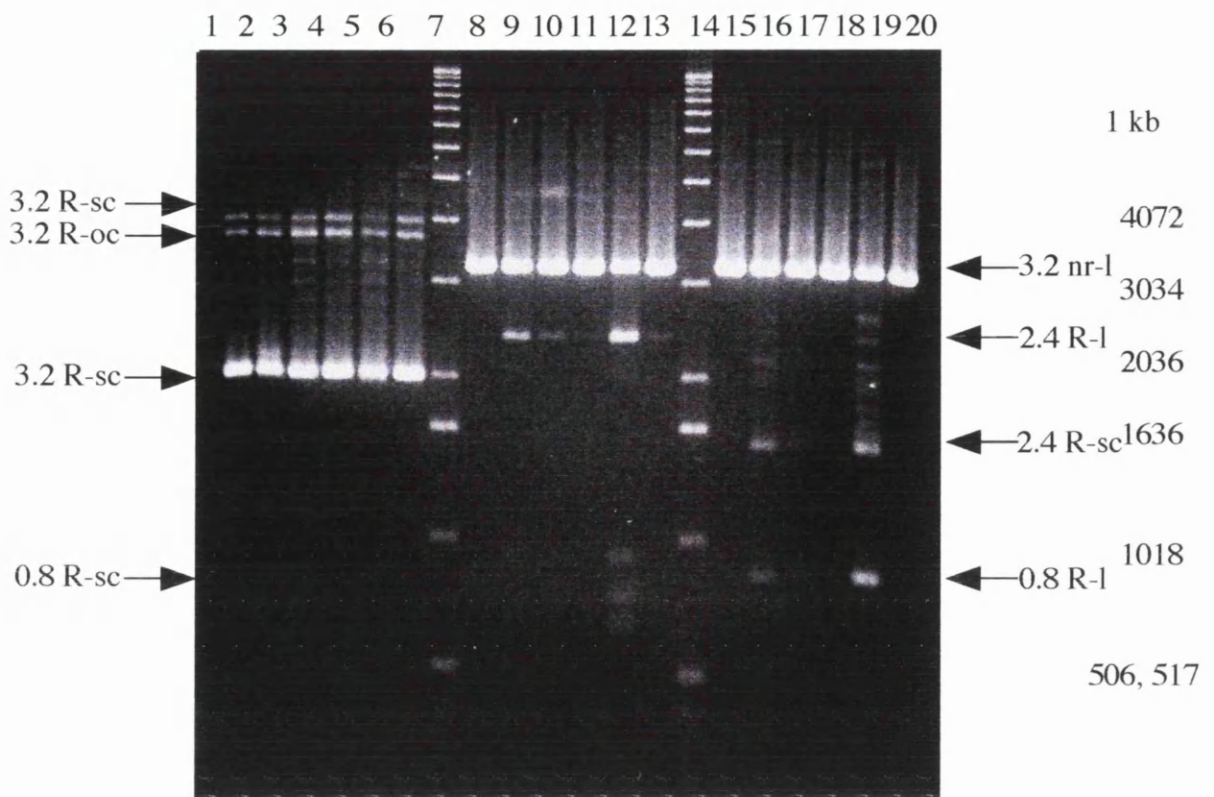
**Figure 6.4 Construction of the recombination substrates**

Diagram showing the cloning steps involved in the construction of the plasmids used in the *EcoRV* experiments. pMSH1 was unsuitable as a substrate because the *res* sites were separated by a length of DNA known to give poor resolution (298 bp).



**Figure 6.5 Resolution of supercoiled substrate pMSH2**

The substrate was designed so that when it is resolved, the chiral phosphorothioate and the radiolabel are both in the smaller recombinant circle. The radiolabelled phosphate is shown as a blue asterisk.

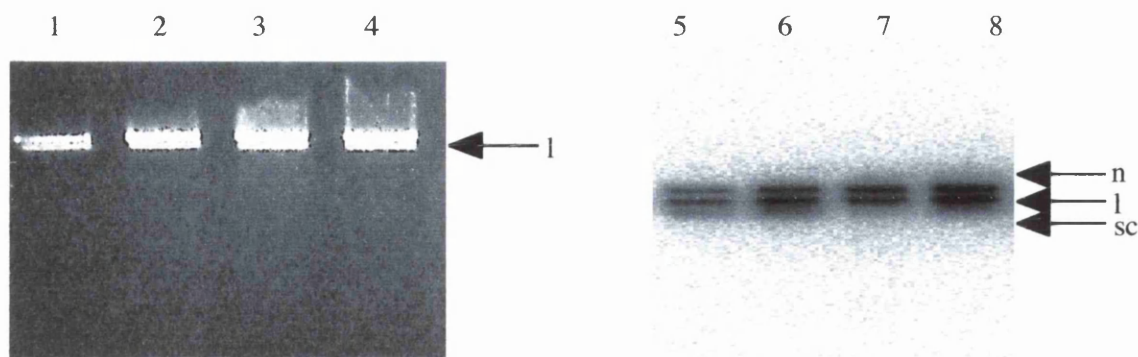


**Figure 6.6 Resolution of pMSH2**

R = recombinant, sc = supercoiled, oc = open circle, l = linear, nr = non recombinant

The reactions were carried out in MD 10 buffer (50 mM glycine/NaOH pH 10, 5 mM EDTA, 5 mM spermidine). Proteins were added to DNA and reacted at 37 °C for 60 minutes. Size markers are shown on the right.

- Lane 1 - pMSH2 + neat wild-type resolvase
- 2 - pMSH2 + 2-fold dilution of wild-type resolvase
- 3 - pMSH2 + 4-fold dilution of wild-type resolvase
- 4 - pMSH2 + 8-fold dilution of wild-type resolvase
- 5 - pMSH2 + neat D102Y
- 6 - pMSH2 + 2-fold dilution of D102Y
- 7 - 1 Kb ladder
- 8 - 13 as lanes 1-6 + *Bgl*II
- 14 - 1 Kb ladder
- 15 - 20 as lanes 1-6 + *Eco*RV



**Figure 6.7 Titration of supercoiled pMSH3 with *EcoRV***

L = linear, n = nicked, sc = supercoiled

Two radiolabelled bands result when the supercoiled substrate is cut with *EcoRV*. The top band represents nicked plasmid only (the strand with  $S_p$  modification). The top strand is linearised and represents the  $R_p$  isomer. The ratio of nicked to linear is 50%, expected from the chiral phosphorus. The unlabelled DNA is fully digested at all concentrations of *EcoRV* (lanes 1-4).

- Lane 1 - pMSH3 + 1 unit of *EcoRV*
- 2 - pMSH3 + 5 units of *EcoRV*
- 3 - pMSH3 + 10 units of *EcoRV*
- 4 - pMSH3 + 20 units of *EcoRV*
- 5 - 8 - phosphoimage of 1-4.

**Table 6.1 Quantitation of *EcoRV* digestion**

	Lane 5A	Lane 5B	Lane 6A	Lane 6B	Lane 7A	Lane 7B	Lane 8A	Lane 8B	Mean $\pm$ SD
nicked	32.57	33.29	33.00	37.96	49.81	49.95	42.35	42.45	40.17 $\pm$ 2.61
linear	56.01	55.29	60.00	51.66	50.19	50.05	44.96	57.55	53.21 $\pm$ 2.13
uncut	11.43	11.72	5.88	10.39	—	—	5.82	6.01	8.54 $\pm$ 1.78

The titration of pMSH3 with *EcoRV* shows that the DNA is cut as expected.

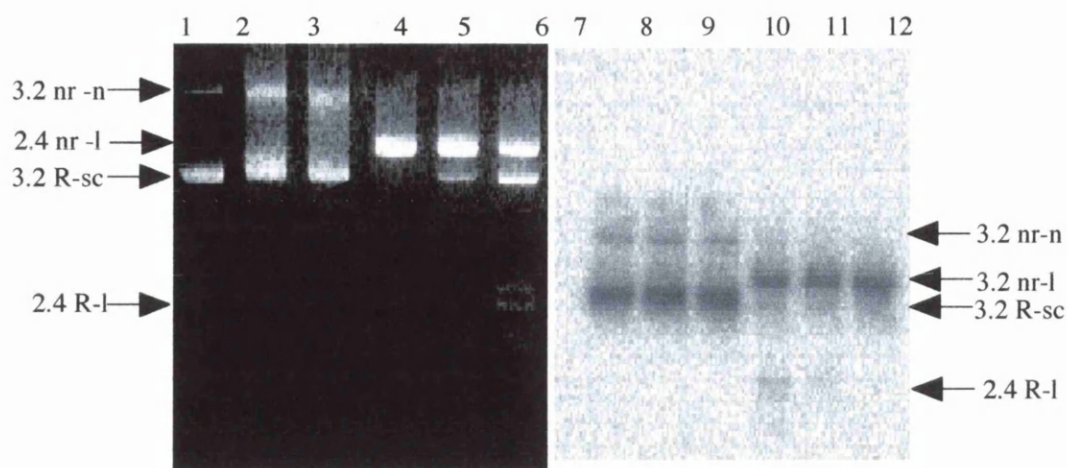
Unlabelled pMSH2 was added to the titration in order to compare the results (see Figure 6.7, lanes 1-4 contain linear pMSH2). The radiolabelled pMSH3, containing the racemic phosphorothioate, gives three products on treatment with *EcoRV* (see Figure 6.7, lanes 5-8).

The R<sub>p</sub>-containing strand and the wild-type strand were cut to completion, yielding the linear product. The more retarded band represents nicked substrate, where the *EcoRV* nicked the S<sub>p</sub> isomer in the wild-type strand. The ratio of linear to nicked is 50:50, as predicted for the racemic phosphorothioate. The third band, below the linear band may represent residual supercoiled substrate, suggesting that the phosphorothioate-containing radiolabelled substrate is cut less efficiently than the natural plasmid.

### 6.5 Resolution of pMSH3

The ratio of diastereoisomers is 50:50, as shown in Figure 6.7. When carrying out resolution reactions on pMSH3, a known quantity of cold non-phosphorothioate pMSH2 was added, so that both the cold and the labelled reaction products could be monitored. All reaction products were run on 1.5% agarose gels (Section 3.21), then dried and analysed with the phosphoimager (Section 3.29.2).

The results of the first resolution assay with pMSH3 are shown in Figure 6.8.



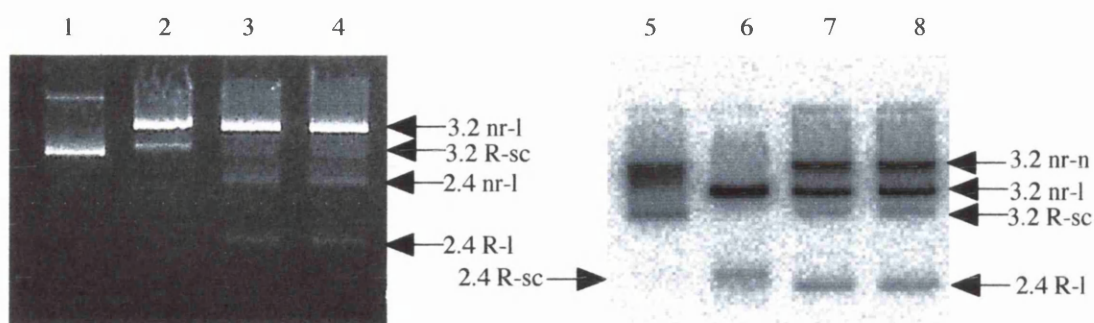
**Figure 6.8 Resolution of pMSH3**

R = recombinant, nr = non recombinant, n = nicked, sc = supercoiled, l = linear

The reactions were carried out in MD 10 buffer (50 mM glycine/NaOH pH 10, 5 mM EDTA, 5 mM spermidine). Proteins were added to DNA and reacted at 37 °C for 60 minutes.

- Lane 1 - pMSH3 + dilution buffer
- 2 - pMSH3 + 2-fold dilution of wild-type resolvase
- 3 - pMSH3 + undiluted D102Y resolvase
- 4 - lane 1 + *Bgl*III
- 5 - lane 2 + *Bgl*III
- 6 - lane 3 + *Bgl*III
- 7-12 - phosphoimage of lane 1-6





**Figure 6.9 Resolution of pMSH3**

R = recombinant, nr = non recombinant, n = nicked, l = linear, sc = supercoiled

The reactions were carried out in MD 10 buffer (50 mM glycine/NaOH pH 10, 5 mM EDTA, 5 mM spermidine). Proteins were added to DNA and reacted at 37 °C for 60 minutes.

- Lane 1 - pMSH3 + undiluted D102Y
- 2 - lane 1 + *Bgl*II
- 3 - lane 1 + 1 unit *Eco*RV
- 4 - lane 1 + 5 units *Eco*RV
- 5 - 8 phosphoimage of lanes 1 - 4

More resolution occurred with the D102Y mutant than with wild-type resolvase. The corresponding phosphoimage shows that the smaller circle remains labelled after resolution. There is almost no resolution product from wild-type resolvase, indicating that the modifications at the crossover site are inhibiting the reaction.

Figure 6.9 shows a resolution reaction under optimized conditions. The 50:50 mixture of nicked to linear DNA in the *EcoRV* digest is very apparent for the resolution substrate. However, only approximately 12% recombination occurred, so it is impossible to predict the source of the linear labelled product. The reaction requires a much higher recombination efficiency before it can be deduced whether the linear or nicked band is depleted, as the amount of the 800 bp linear product increases. This would correlate directly with the preference of Tn3 resolvase for R<sub>p</sub> or S<sub>p</sub> phosphorothioates.

## 6.6 Discussion

All attempts to improve the efficiency of the recombination of the modified substrate pMSH3 were unsuccessful. Increasing the concentration of enzyme and reaction times did not improve recombinational efficiency. It was concluded that the phosphorothioate at the crossover site is severely inhibiting the recombination reaction, and that any recombination products observed were from a small portion of pMSH3 that did not contain any phosphorothioate (i. e. had a natural phosphodiester at the point of cleavage). The inefficiency of the reaction was probably due to a combination of the mutations in both *res* sites in the substrate, and the phosphorothioate modification. The *EcoRV* site in the resolution substrates alone does not account for the inefficiency, since the cold substrate pMSH2 resolved fairly well with wild-type resolvase. The introduction of both mutated *res* site Is in pMSH2, gave no recombination with wild-type resolvase, but the substrate was resolved with D102Y. However, the phosphorothioate modification in pMSH3 caused a decrease in the efficiency of recombination, even with D102Y. Both R<sub>p</sub> and S<sub>p</sub> configurations of the phosphorothioate inhibit

recombination. The strategy of using *EcoRV* to determine the phosphorothioate stereochemistry is therefore impracticable.

The experiment might be feasible with wild-type substrate with a phosphorothioate-modified oligonucleotide (i.e. no *EcoRV*) sites. A wild-type substrate e.g. pMM5 could be modified with a wild-type *SstI-EcoRI* site (<sup>32</sup>P-labelled as before). After resolution the recombinant product, containing the chiral phosphorothioate could be cut with *PsiI* and radiolabelled. *PsiI* cuts at the centre of site I sequence TTA↓ATT. Following purification of the products an enzymatic resolution could be carried out. The nuclease P1 cleaves S<sub>p</sub> diastereoisomers of phosphorothioates only and snake venom phosphodiesterase cleaves R<sub>p</sub> diastereoisomers of phosphorothioates only. If the purified recombinant product cuts with either of these nucleases and it should be possible to determine which isomer was preferentially hydrolysed.

## 6.7 Future Experiments

Another possible approach to solve this complex problem would be to purify chiral phosphorothioate oligonucleotides using HPLC on a chiral column. The study of the chirality of the strand transfer step of phage Mu used 24-mer oligonucleotides (Mizuuchi and Adzuma, 1991). This strategy proved to be too problematic in this work. The phosphorothioate modification is at the centre of a 48-mer oligonucleotide. Separation of the stereoisomers was therefore very difficult. The columns were run at elevated temperatures, but the length of the oligonucleotide was such that the subtle differences in chirality were too small to detect.

Another approach would be to make the target oligonucleotide out of two shorter pieces. T T A ↓ T(S)A T. The oligo with the phosphorothioate at the 5' end could be separated by the reverse phase method using the trityl group (Mizuuchi and Adzuma, 1991). Alternatively, the deprotected form could be purified (Grasby and Connolly, 1988). The two pieces would then be religated. This could be done

after the  $^{32}\text{P}$  labelling in the final synthetic step, thereby making supercoiled plasmid. Or it could be religated as the oligonucleotide, then the religated product could be purified.

If *EcoRV* is good at discriminating between  $\text{R}_p$  and  $\text{S}_p$  phosphorothioates, another way to get one stereoisomer would be to cut the double stranded oligonucleotide with *EcoRV*, then purify the uncut DNA. However, only one stereoisomer can be obtained in this way.

**Chapter 7: Further Work**

## 7.1 Future Perspectives

Two different approaches were used to study the strand exchange mechanism of Tn3 resolvase. The crosslinking approach, would allow us to differentiate between different subunits of protein that bind to the three different binding sites, by "fixing" a particular subunit to its site. The introduction of a chiral DNA substrate at the point of protein-induced cleavage was also used to attempt to determine the number of catalytic steps involved in the reaction.

Site-specific crosslinking to Tn3 resolvase using the disulphide approach proved to be of limited usefulness. The thiol groups that were introduced, as novel cysteine mutants of Tn3 resolvase, gave proteins that had similar binding and resolution activity to wild-type protein. However difficulties arose with the thiol-containing oligonucleotides and their reactivity. The (protected) thiol-containing phosphoramidite that was synthesised was predicted to be reactive enough, although chemical degradation might have resulted in less free thiol being available to be crosslinked. On the other hand, the mercaptomethyl compound was far too reactive, and resulted in the kinetic product of the disulfide (Section 2.11). An alternative to this would be to use a thiol that was separated from the pyrimidine ring by a two carbon chain. This should lead to a reagent of intermediate reactivity which may be more able to form disulfides with the cysteine thiols. The chemical synthesis that led to the mercaptomethyl analogue (Figure 2.8.3) *via* palladium cross coupling reaction could also be carried out. It has been reported that 5-mercaptomethyl uracil was coupled to the sugar portion, yielding the desired product. In future this route could be tried, as an alternative to the chemistry described in this thesis.

**Bibliography**

- S. S. Abdel-Meguid, N. D. F. Grindley, N. S. Templeton and T. A. Steitz, 1984, *Proc. Natl. Acad. Sci. USA*, **81**, 2001-2005.
- M. Ahmadian and D. E. Bergstrom, 1998, *Nucleosides, Nucleotides*, **17**, 1183-1190.
- B. Angus and R. G. R. Bacon, 1958, *J. Chem. Soc.*, 774-784.
- P. H. Arnold, D. G. Blake, N. D. F. Grindley, M. R. Boocock and W. M. Stark, 1999, *EMBO J.*, **18**, 1407-1414.
- P. Arnold, 1997, PhD thesis
- R. W. Armstrong, S. Gupta and F. Wheliam, 1989, *Tetrahedron Lett.*, **30**, 2057-2060.
- B. Barbier, M. Charlier and J-C. Maurizot, 1984, *Biochemistry*, **23**, 2933-2939.
- T. J. Bardos, M. P. Kotick and C. Szantay, 1966, *Tett. Lett.*, **16**, 1759-1764.
- T. J. Bardos, R. R. Herr and T. Enkoji, 1955, *J. Am. Chem. Soc.*, **77**, 960-963.
- B. Bartholomew, G. A. Kassavetis, B. R. Braun and E. P. Geiduschek, 1990, *EMBO J.*, **9**, 2197-2205.
- B. Bartholomew, B. R. Braun, G. A. Kassavetis, and E. P. Geiduschek, 1994, *J. Biol. Chem.*, **269**, 18090-18095.
- S. L. Beaucage and R. P. Iyer, 1993, *Tetrahedron*, **49**, 1925-1963.
- H. W. Benjamin and N. R. Cozzarelli, 1986, In proceedings of the Robert A. welch Foundation Conferences on Chemical Research, (Welch Foundation, Houston), p107-126.
- A. L. Bednarz, M. R. Boocock and D. J. Sherratt, 1990, *Genes Dev.*, **4**, 2366-2375.
- D. E. Bergstrom, P. Beal, J. Jenson and X. Lin, 1991, *J. Org. Chem.*, **56**, 5598-5602.
- B. H. C. Birnhoim and J. Doly, 1979, *Nucleic Acids Res.*, **6**, 1513-1523.
- D. G. Blake, 1993, PhD thesis.
- D. G. Blake, M. R. Boocock, D. J. Sherratt and W. M. Stark, 1995, *Current Biol.*, **5**, 1036-1046.
- E. E. Blatter, Y. W. Ebright and R. E. Ebright, 1992, *Nature*, **359**, 650-652.



- C. Bleasdale and S. B. Ellwood and B. T. Golding, 1990, *J. Chem. Soc., Perkin Trans. I*, 803-804.
- J. B. Bliska, H. W. Benjamin and N. R. Cozzarelli, 1991, *J. Biol. Chem.*, **226**, 2041-2047.
- M. R. Boocock, L. J. Brown and D. J. Sherratt, 1987, In *DNA Replication and Recombination*, T. J. Kelly and R. McMacken, eds. (Alan R. Liss, New York), pp703-718.
- M. R. Boocock, X. Zhu and N. D. F. Grindley, 1995, *EMBO J.*, **14**, 5129-5140.
- F. G. Bordwell and H. M. Andersen, 1953, *J. Amer. Chem. Soc.*, **75**, 6019-6022.
- D. H. Bradley and M. M. Hanna, 1992, *Tetrahedron Lett.*, **33**, 6223-6226.
- J. L. Brown, 1986, PhD thesis.
- L. Campanini, A. Dureault and J. C. Depezay, 1995, *Tetrahedron Lett.*, **36**, 8015-8018.
- G. Chang and M. P. Mertes, *J. Org. Chem.*, **52**, 3625-3631.
- Y. Chen and R. H. Ebright, 1993, *J. Mol. Biol.*, **230**, 453-460.
- B. C. F. Chu and L. E. Orgel, 1988, *Nucleic Acids Res.*, **16**, 3671-3691.
- B. C. F. Chu and L. E. Orgel, 1992, *Nucleic Acids Res.*, **20**, 2497-2502.
- R. F. R. Church and M. J. Weiss, 1970, *J. Org. Chem.*, **35**, 2465-2471.
- M. A. Cleary, P. S. Pendergrast and W. Herr, 1997, *Proc. Natl. Acad. Sci. USA*, **94**, 8450-8455.
- R. E. Cline, R. M. Fink and K. Fink, 1959, *J. Am. Chem. Soc.*, **81**, 2521-2527.
- B. A. Connolly, B. V. L. Potter, F. Eckstein, A. Pingoud and L. Grotjahn, 1984, *Biochemistry*, **23**, 3443-3453.
- M. Cowart, S. J. Benkovic and H. A. Nash, 1991, *J. Mol. Biol.*, **220**, 621-629.
- F. J. Dinan, J. Chodkowski, J. P. Barren, D. M. Robinson and D. V. Reinhardt, 1982, *J. Org. Chem.*, **47**, 1769-1772.
- Q. Dong, E. E. Blatter, Y. W. Ebright, K. Bister and R. H. Ebright, 1994, *EMBO J.*, **13**, 200-204.
- P. Dumoulin, P. Oertel-Buchheit, M. Granger-Schnarr and M. Schnarr, 1993, *Proc. Natl. Acad. Sci. USA*, **90**, 2030-2034.

- P. Dumoulin, R. H. Ebright, R. Knegtel, R. Kaptein, M. Granger-Schnarr and M. Schnarr, 1996, *Biochem.*, **35**, 4279-4286.
- Y. Endo, K. Shudo and T. Okamoto, 1982, *J. Am. Chem. Soc.*, **104**, 6393-6397.
- A. Engelman, K. Mizuuchi and R. Craigie, 1991, *Cell*, **67**, 1211-1121.
- H. A. Flaschka, A. J. Barnard and P. E. Sturrock, 1969, in *Quantitative Analytical Chemistry, Short Introduction to Practice*, Barnes & Noble, Inc., New York
- R. Gill, F. Heffron and S. Falkow, 1979, *Nature*, **282**, 797-801.
- A. Giner-Sorolla and L. Medrek, 1966, *J. Med. Chem.*, **9**, 97-101.
- A. Giner-Sorolla and L. Medrek, 1978, in "Nucleic Acid Chemistry" Part 1, L. B. Townsend and R. S. Tipson, Eds, pp. 83-85.
- G. D. Glick, 1991, *J. Org. Chem.*, 1991, **56**, 6746-6747.
- J. Goodchild, 1990, *Bioconj. Chem.*, **1**, 165-187.
- J. T. Goodwin and G. D. Glick, 1993, *Tett. Lett.*, **35**, 5549-5552.
- J. A. Grasby and B. A. Connolly, 1992, *Biochemistry*, **31**, 7855-7861.
- N. D. F. Grindley, 1993, *Science*, **262**, 738-740.
- N. D. F. Grindley, 1994, In *Nucleic Acids and Molecular biology*, Vol 8, eds. F. Eckstein and D. M. J. Lilley pp. 236-267.
- E. S. Gruff and L. E. Orgel, 1991, *Nucleic Acids Res.*, **19**, 6849-6854.
- V. S. Gupta, G. L. Bubbar, J. B. Meldrum and J. R. Saunders, 1975, *J. Med. Chem.*, **18**, 973-976.
- M. M. Hanna, S. Dissinger, B. D. Williams and J. E. Colston, 1989, *Biochemistry*, **28**, 5814-5820.
- M. M. Hanna, Y. Zhang, J. C. Reidling, M. J. Thomas and J. Jou, 1993, *Nucleic Acids Res.*, **21**, 2073-2079.
- M. E. Harris, J. M. Nolan, A. Malhotra, J. W. Brown, S. C. Harvey and N. R. Pace, 1994, *EMBO J.*, **13**, 3953-3963.
- G. F. Hatfull and N. D. F. Grindley, 1986, *Proc. Natl. Acad. Sci. USA*, **83**, 5429-5433.
- G. F. Hatfull and N. D. F. Grindley, 1988, In *genetic Recombination*, R. Kucherlapati and G. R. Smith, eds. (American Society for Microbiology, Washington, DC), p357-396

- M. E. Harris, J. M. Nolan, A. Malhotra, J. W. Brown, S. C. Harvey and N. R. Pace, 1994, *EMBO J.*, **13**, 3953-3963.
- H. Hayakawa, H. Tanaka, K. Obi, M. Itoh and T. Miyasaka, 1987, *Tetrahedron Lett.*, **28**, 87-90.
- B. He, D. L. Riggs and M. M. Hanna, 1995, *Nucleic Acids Res.*, **23**, 1231-1238.
- F. Heffron, B. J. McCarthy, H. Ohtsubo and E. Ohtsubo, *Cell*, **18**, 1153-1163.
- K. Hideg, C. P. Sar, O. H. Hankovszky, T. Tamas and G. Jerkovich, 1992, *Synthesis*, 390-394.
- H. Huang, R. Chopra, G. L. Verdine and S. C. Harrison, 1998, *Science*, **282**, 1669-1675.
- R. E. Hughes, G. E. Hatfull, P. Rice, T. A. Steitz and N. D. F. Grindley, 1990, *Cell*, **63**, 1331-1338.
- R. E. Hughes, P. Rice, T. A. Steitz and N. D. F. Grindley, 1993, *EMBO J.*, **12**, 1447-1458.
- K. B. Jensen, B. L. Atkinson, M. C. Willis, T. H. Koch and L. Gold, 1995, *Proc. Natl. Acad. Sci. USA*, **92**, 12220-12224.
- M. Katouzian-Safadi, B. Laine, F. Chartier, J. Cremet, D. Belaiche, P. Sautiere and M. Charlier, 1992, *Nucleic Acids Res.*, **19**, 4937-4941.
- A. K. Kennedy, D. B. Haniford and K. Mizuuchi, 2000, *Cell*, **101**, 295-305.
- E. Kilbride, M. R. Boocock and W. M. Stark, 1999, *J. Mol. Biol.*, 1999.
- L. Kimzey and W. S. Dynan, 1998, *J. Biol. Chem.*, **273**, 13768-13775.
- P. A. Kitts and H.A. Nash (a), 1988, *Nucl.Acids.Res.*, **16**, 6839-6856.
- P. A. Kitts and H. A. Nash (b), 1988, *J. Mol. Biol.*, 1988, **204**, 95-107.
- N. Kharasch, W. King and T. C. Bruice, 1955, *J. Am. Chem. Soc.*, **77**, 931-934.
- A. Klippel, G. Mertens, T. Patschinsky and R. Kahmann, *EMBO J.*, **7**, 1229-1237.
- M. P. Kotick, C. Szantay and T. J. Bardos, 1969, *J. Org. Chem.*, **34**, 3806-3813.
- M. Kulka, 1956, *Can. J. Chem.*, **34**, 1093-1100.
- M. R. Kurpiewski, M. Koziolkiewicz, A. Wilk, W. J. Stec and L. Jen-Jacobson, *Biochemistry*, **35**, 8846-8854.

- T. Lagrange, T. K. Kim, G. Orphanides, Y. W. Ebright, R. H. Ebright and D. Reinberg, 1996, *Proc. Natl. Acad. Sci. USA*, **93**, 10620-10625.
- P. Langen and D. Barwolff, 1975, *Acta biol. med. germ.*, **34**, K7-K10 and 1975, German Patent No. 114949.
- K. Lee and J. Hayes, 1997, *Proc. Natl. Acad. Sci. USA*, **94**, 8959-8964.
- C. H. Lee and Y. H. Kim, 1991, *Tetrahedron Lett.*, **32**, 2401-2404.
- J. Leonard, B. Lygo and G. Procter, 1989, In *Advanced Practical Organic Chemistry*, Stanley Thornes (pubs)pp. 69-74.
- A. E. Leschziner, M. R. Boocock and N. D. F. Grindley, 1995, *J. Mol. Microbiol.*, **15**, 865-870.
- T. S. Lin and Y. S. Gao, 1983, *J. Med. Chem.*, **26**, 598-601.
- T. Lui, E. F. DeRose and G. P. Mullen, 1994, *Protein Science*, **3**, 1286-1295.
- T. Lui, D. Lui, E. F. DeRose and G. P. Mullen, 1993, *J. Biol. Chem.*, **268**, 16309-16315.
- J. Liu, M. Sodeoka, W. S. Lane and G. L. Verdine, 1994, *Proc. Natl. Acad. Sci. USA*, **91**, 908-912.
- M. T. H. Lui, 1986, In *Chemistry of Diazirines Vol.II*, CRC Press Inc. Boca Raton Florida, Florida pp. 77-81.
- A. MacDonald, 1999, PhD thesis.
- A. M. MacMillan and G. L. Verdine, 1991, *Tetrahedron*, **47**, 2603-2616.
- V. A. Malkov, I. Biswas, R. D. Camerini-Otero and P. Hsieh, 1997, *J. Biol. Chem.*, **272**, 23811-23817.
- D. J. Martin and C. C. Greco, 1968, *J. Chem. Soc.*, 1275-1276.
- A. Marquet, F. Frappier, G. Guillermin, M. Azoulay, D. Florentin and J. C. Tabet, 1993, **115**, 2139-2145.
- J. Matulic-Adamic and K. A. Watanabe, 1988, *Chem. Pharm. Bull.*, **36**, 1554-1557.
- J. M. Mazzairelli, M. R. Ermacora, R. O. Fox and N. D. F. Grindley, 1993, *Biochemistry*, **32**, 2979-2986.
- J. D. McGhee and P. H. von Hippel, 1975, *Biochemistry*, **14**, 1281-1285.

- McGregor, M. V. Rao, G. Duckworth, P. G. Stockley and B. A. Connolly, 1996, *Nucleic Acids Res.*, **24**, 3173-3180.
- M. J. McIlwraith, 1995, PhD thesis.
- M. J. McIlwraith, M. R. Boocock and W. M. Stark, 1996, *J. Mol. Biol.*, **260**, 299-303.
- M. J. McIlwraith, M. R. Boocock and W. M. Stark, 1997, *J. Mol. Biol.*, **266**, 106-121.
- K. M. Meisenheimer and T. H. Koch, 1997, *Crit. Rev. Biochem. Mol. Biol.*, **32**, 101-140.
- M. J. Melnick and S. M. Weinreb, *J. Org. Chem.*, 1988, **53**, 850-854.
- K. L. Meyer and M. M. Hanna, 1996, *Bioconj. Chem.*, **7**, 401-412.
- A. M. Michelson and A. R. Todd, 1955, *J. Chem. Soc.*, 816-823.
- K. Mizuuchi and K. Adzuma, 1991, *Cell*, **66**, 129-140.
- L. L. Murley and N. D. F. Grindley, 1998, *Cell*, **95**, 553-562.
- T. Nagamachi, P. F. Torrence, J. A. Waters and B. Witkop, 1972, *J. Chem. Soc., Chem. Comm.*, 1025-1026.
- T. Nagamachi, J. L. Fourrey, P. F. Torrence, J. A. Waters and B. Witkop, 1974, *J. Med. Chem.*, **17**, 403-406.
- N. Naryshkin, A. Revyakin, Y. Kim, V. Mekier and R. H. Ebright, 2000, *Cell*, **101**, 601-611.
- T. T. Nikiforov and B. A. Connolly, 1992, *Nucleic Acids Res.*, **20**, 1209-1214.
- K. K. Ogilvie, S. L. Beaucage, A. L. Schiffman, N. Y. Theriault and K. L. Sadana, 1978, *Can. J. Chem.*, **56**, 2768-2780.
- G. Pashev, S. I. Dimitrov and D. Angelov, 1991, *Trend. Biochem. Sci.*, **16**, 323-326.
- P. S. Pendergrast, Y. Chen, Y. W. Ebright and R. E. Ebright, 1992, *Proc. Natl. Acad. Sci. USA*, **89**, 10287-10291.
- R. R. Reed and N. D. F. Grindley, 1981, *Cell*, **25**, 721-728.
- R. R. Reed and C. D. Moser, 1984, *Cold Spring Harbour Symp. Quant. Biol.*, **49**, 245-249.
- V. Rimphanitchayakit and N. D. F. Grindley, 1990, *EMBO J.*, **9**, 719-725.

- V. Rimphanitchayakit, G. F. Hatfull and N. D. F. Grindley, 1989, *Nucleic Acids Res.*, **17**, 1035-1050.
- A. Romieu, S. Bellon, D. Gasparutto and J. Cadet, *Organic Lett.*, **2**, 1085-1088.
- M. R. Sanderson, P. S. Rice, P. A. Goldman, G. F. Hatfull, N. D. F. Grindley and T. A. Steitz, 1990, *Cell*, **63**, 1323-1329.
- Y. S. Sanghvi, Z. Guo, H. M. Pfundheller and A. Converso, 2000, *Org. Process Res. & Dev.*, **4**, 175-181.
- J. Sekiguchi and S. Shuman, 1996, *EMBO J.*, **15**, 3448-3457.
- P. V. Sergiev, I. N. Lavrik, V. A. Wlasoff, S. S. Dokudovskaya, O. A. Dontsova, A. A. Bogdanov and R. Brimacombe, 1997, *RNA*, **3**, 464-475.
- Y. F. Shealy, C. A. O'Dell, G. Arnett and W. M. Shannon, 1986, *J. Med. Chem.*, **29**, 79-84.
- D. Sherratt, 1988, In *Mobile DNA*, D. E. Berg and H. Howe, eds. (American Society for Microbiology, Washington, D.C), pp163-184.
- G. T. Shiau, R. F. Schinazi, M. S. Chen and W. H. Prusoff, 1980, *J. Med. Chem.*, **23**, 127-133.
- D. Shore, J. Langowski and R. L. Baldwin, 1981, *Proc. Natl. Acad. Sci. USA*, **78**, 4833-4837.
- N. D. Sinha, J. Biernat and H. Koster, 1983, *Tetrahedron Lett.*, **24**, 5843-5846.
- W. M. Stark, D. J. Sherratt and M. R. Boocock, 1989, *Cell*, **58**, 779-790.
- W. M. Stark and M. R. Boocock and D. J. Sherratt, 1989, *Trend Genet.*, **5**, 304-309.
- W. M. Stark, N. D. F. Grindley, G. F. Hatfull and M. R. Boocock, 1991, *EMBO J.*, **10**, 3541-3548.
- W. M. Stark, M. R. Boocock and D. J. Sherratt, 1992, *Trends in Genetics*, **8**, 432-439.
- W. M. Stark and M. R. Boocock, 1995, In *Mobile Genetic Elements* (Frontiers in Mol. Biology Series), D. J. Sherratt, ed., (Oxford University Press), p101-129.
- F. W. Studier, A. H. Rosenberg, J. J. Dunn and J. W. Dubendorff, 1990, *Methods Enzym.*, **185**, 60-89.
- J. G. Sutcliffe, 1978, *Cold Spring Harbour Symp. Quant. Biol.*, **43**, 77-90.

- J. J. Tate, J. Persinger and B. Bartholomew, 1998, *Nucleic Acids Res.*, **26**, 1421-1426.
- L. A. Taylor and R. E. Rose, 1988, *Nucleic Acids Res.*, **16**, 358-360.
- B. H. Thorogood, J. A. Grasby and B. A. Connolly, 1996, *J.Biol.Chem.*, 1996, **271**, 8855-8862.
- R. L. Tinker, G. A. Kassavetis and E. P. Geiduschek, *EMBO J.* **13**, 5330-5337.
- A. N. Topin, O. M. Gritsenko, M. G. Brevnov, E. S. Gromova and G. A. Korschunova, 1998, *Nucleotid. Nucleosid.*, **17**, 1163-1175.
- P. F. Torrence, J. A. Waters and B. Witkop, 1978, in "Nucleic Acid Chemistry" Part 1, L. B. Townsend and R. S. Tipson, Eds, pp. 367-370.
- H. Urlaub, V. Kryft, O. Bischof, E. Muller and B. Wittmenn-Liebold, 1995, *EMBO J.*, **14**, 4578-4588.
- G. G. Urquhart, J. W. Gates and R. Connor, 1955, *Org. Synth.* , **3**, 363-365.
- S. Verma and F. Eckstein, 1998, *Annu. Rev. Biochem.*, **67**, 99-134.
- M. A. Watson, 1994 Ph.D thesis
- M. A. Watson, M. R. Boocock and W. M. Stark, 1996, *J. Mol. Biol.*, **257**, 317-329.
- S. Weinswieser, 2001, PhD thesis
- J. Welsh and C. R. Cantor, *Trends Biol. Sci.*, 1984, 505-508.
- D. L. Wong, J. G. Pavlovich and N. O. Reich, 1998, *Nucleic Acids Res.*, **26**, 645-649.
- K. R. Williams and W. H. Konigsberg, 1991, *Method. Enzymology*, **208**, 516-539.
- M. C. Willis, B. J. Hicke, O. C. Unhelbeck, T. R. Cech and T. H. Koch, 1993, *Science*, **262**, 1255-1256.
- M. C. Willis, K. A. LeCuyer, K. M. Meisenheimer, O. C. Uhlenbeck and T. H. Koch, 1994, *Nucleic Acids Res.*, **22**, 4947-4952.
- L. J. Wilson, M. W. Hager, Y. A. El-Kattan and D. C. Liotta, *Synthesis*, 1465-1479.
- Y. Z. Xu, Q. Zheng and P. F. Swann, 1992, *Tett. Lett.*, **33**, 5837-5840 and *Tetrahedron*, **48**, 1729-1740.
- H. Yang and H. A. Nash, 1994, *Proc. Natl. Acad. Sci. USA*, **91**, 12183-12187.

- W. Yang and T. A. Steitz, 1995, *Cell*, **82**, 193-207.
- C. Yanisch-Perron, J. Vieira and J. Messing, 1985, *Gene*, **33**, 103-119.
- E. Yuriev, D. Scott and M. M. Hanna, 2000, *J. Mol. Recognit.*, **12**, 337-345
- X. Zhao and D. W. Landry, 1993, *Tetrahedron*, **49**, 363-368.

